

White heat — at last

Britain's political leadership, for the first time in decades, is well placed to take a reasoned, strategic approach to the long-term development of science and innovation. But watch out for resurgent bureaucracy.

For fifty years, British governments have sought to reverse the actual and perceived decline of the nation's relative economic and political clout. The decline — which economic historians date from the zenith of British imperial power at around 1870 — has frequently been attributed to a deep-seated, almost cultural, national aversion to science and technology.

This malaise was articulated most forcefully and famously by C. P. Snow, who in 1959 described the “two cultures” in Britain of the scientist and non-scientist, and argued persuasively that the latter always had the upper hand. Snow's diagnosis was accepted by, among others, Harold Wilson, a technocrat who was elected as Labour prime minister after pledging to forge a new Britain in the “white heat” of scientific and technological revolution.

But Wilson's four governments, like others before and since, were unable or unwilling to get a real handle on the issue of science and innovation. In a country torn by economic and political crises, as Britain was for much of the time from the 1960s to the 1980s, prime ministers and chancellors had other fish to fry. Economic solvency and industrial peace were always their elusive goals; modernization of the science base would have to wait.

Since its nadir in the 1970s, there is no question that a spirit of confidence has rebounded, and science and technology with it. This can be attributed at least in part to Margaret Thatcher, a trained chemist whose loathing for the British academic class was reciprocated to an unhealthy degree when she was prime minister. Her confrontational approach was necessary, however, to shake Britain's universities and laboratories from the complacency in which many of them had wallowed for a large part of the twentieth century.

Now, Prime Minister Tony Blair and his friend, rival and putative successor, Chancellor Gordon Brown, are united in their genuine commitment to science and innovation. A change in government is not anticipated in elections expected next year, but even if one occurs, the current government's level of commitment to science could and should be maintained.

Concerted action

The latest manifestation of that commitment, following a series of important reviews into key aspects of science and technology policy, is a consultation document on a ten-year investment strategy for science and innovation, launched last week by Brown and fellow ministers (see http://www.hm-treasury.gov.uk/consultations_and_legislation/science_innov/consult_sciinnov_index.cfm). The source of the document is significant because so much real fiscal power in Britain rests not with elected representatives, or even with the great departments of state, but rather with the mandarins of Her Majesty's Treasury. Snow regarded high-handed Treasury officials — traditionally educated in politics, philosophy and economics at Oxford University — as the epitome of the crisis that he diagnosed. The significance of Brown's insistence on their commitment to science and innovation cannot be understated.

The document, the exact financial implications of which will not be clear until later in the year, makes a good effort to identify the key questions and problems that need to be tackled in the British

research system. Its authors deserve the congratulations of the scientific community for setting out these issues and proposing an assault on them based not on a few flashy initiatives, but on concerted action over a realistic period of time.

But there is also cause for alarm, both in what is discussed in the document and in what isn't. It emphasizes enhanced links between university and industry. The assumption that more industrial involvement in university departments is a ‘good thing’ was certainly valid twenty years ago. A lot has changed since then, however, and the industrial links enjoyed by strong departments at British universities are now quite solid.

More movement is required by those in business than by the academics: with a few exceptions (primarily in the pharmaceutical industry), British industrial companies are amateurish and lax in their approach to research and development. The government's tax incentives for business research and development are praiseworthy. But an insistence that universities do even more to address economic problems could undermine the academic creativity that history proves to be a significant source of profound technological innovation.

Bureaucratic burden

Alarm bells should also ring in response to the document's suggestion that universities obtain full returns on their costs “at the project level”. The idea that they can plan a sustainable future when faculty staff and other core costs depend on individual research grants is a nonsense. This is also a recipe for yet more bureaucracy. It is essential that the system for funding British universities, currently being reformed by the government, reduces the burden of assessment endured by academics.

The document also contains significant omissions that must be corrected during the consultation period. In particular, it contains no mention of the needs of the developing world, although this government has consistently invoked the state of the South as the scandal of our age. Both altruistically and for enlightened self-interest, investment in building the capacity for research in Asia and Africa has to feature in a ten-year plan for UK science. This is all the more urgent given the government's failure so far to recognize the value of such investment in its aid programmes.

Interested parties are invited to comment on Brown's document by 30 April. Individuals, as well as their laboratories, universities and professional societies, should make their voices heard in what could be an important process in the steady rebuilding of British science and innovation, and national self-confidence in both.

Scientists and engineers of a reflective bent will be delighted that the Treasury is leading the development and promotion of a thorough, long-term plan for science and innovation. The initiative is a measure of how far British science and innovation has come since Snow's diagnosis. Whether industry will rise to the opportunity this presents, and whether the government will take it forward with the appropriate lightness of touch, are important questions that remain to be answered. But for the first time in decades, there are reasonable grounds for optimism. ■





Pet theory

Links investigated between allergies and kids' contact with cats
p354



Split over fusion

No resolution in clash over site for ITER fusion project
p355



Getting serious?

Japanese science council hopes for increased clout
p357



Windows on ET

Microsoft founder funds alien-hunting observatory
p358

NASA seeks robotic rescuers to give Hubble extra lease of life

Tony Reichhardt, Washington

NASA is seeking to damp a firestorm over its decision to stop astronaut servicing of the Hubble Space Telescope by lining up ideas for robots that could do the job.

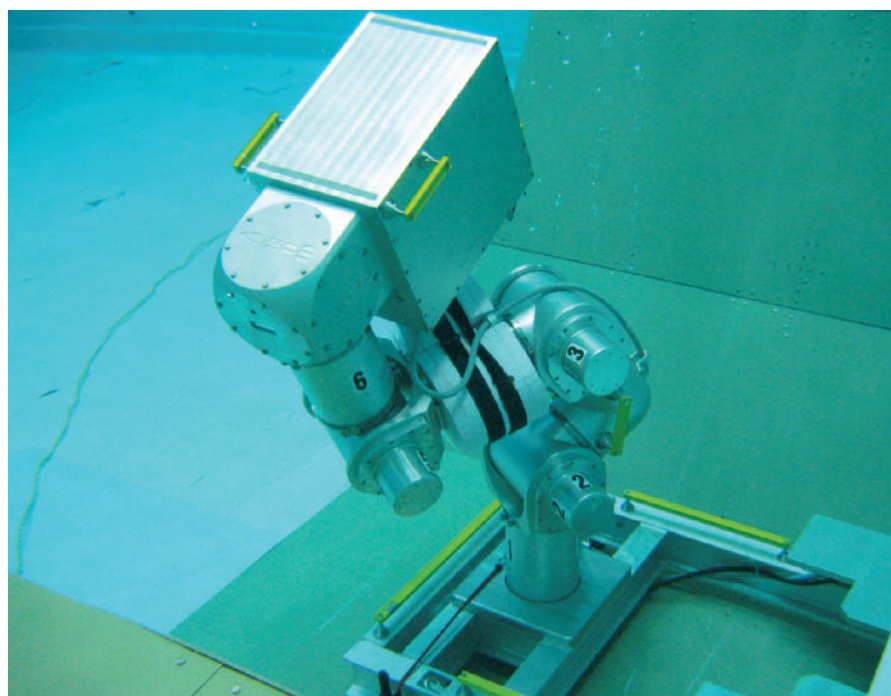
Dozens of teams from industry bodies and academic institutions have responded to an agency request for schemes to use a robotic spacecraft to replace the telescope's batteries and gyroscopes and install two new science instruments.

These were traditionally jobs for astronauts. But NASA science chief Ed Weiler told the National Academies of Sciences' Space Studies Board last week that "it's not a big leap of faith" to imagine that an automated mission could extend the Hubble's working life. A robot could change the telescope's batteries, which are expected to run out of power around 2007.

The deadline for submitting ideas was 22 March. NASA has not yet issued a formal request for bids on such a mission, but Hubble programme executive Michael Moore told the board: "We're actually getting ideas from folks who can do things."

The academy is also forming a committee to study how to service the telescope and extend its life. NASA administrator Sean O'Keefe has vowed not to send astronauts back to Hubble owing to safety concerns with the space shuttle — although critics charge that budgetary and political considerations have also swayed the decision. Shuttle crews have upgraded the telescope four times since its 1990 launch.

Among the teams proposing a robotic mission is one from Lockheed Martin, which built Hubble. At least three systems in development could do the job, says Bruce McCandless, a former astronaut who helped design the procedures for servicing the telescope and is now chief scientist for reusable space transportation systems at Lockheed Martin in Denver. One is made by Canadian company MD Robotics of Brampton, Ontario, which also made the robot arms on the space shuttle and space station. Another is the anthropomorphic Robonaut under study at NASA's Johnson Space Center in Houston. The third is being built by the Uni-



Helping hand? The University of Maryland's robotic arm is one contender for a Hubble mission.

versity of Maryland's Space Systems Laboratory in College Park, a partner in the Lockheed Martin proposal.

The Maryland system, which includes robotic arms bearing tools like those used by astronauts, has been tested in the laboratory and in underwater tanks to simulate weightlessness. It was intended for a shuttle test flight this year, but NASA cut the test's funding when the space station hit financial trouble several years ago. "I've got pieces for 70% of a flight robot locked up in my lab right now," says David Akin, who leads the Maryland project.

Using robots with astronauts on a Hubble repair would not be a big technical stretch, Akin and McCandless say. Having a robot do the whole job is more ambitious and would take longer, says Akin, but is still within reach. And NASA could decide to scrap the more challenging tasks, such as adding new instruments.

Another team, led by SkyCorp of Huntsville, Alabama, has suggested sending up a

robot mission to attach a low-thrust rocket to Hubble and bring it down to the International Space Station for servicing by astronauts. But Steven Beckwith, head of the Space Telescope Science Institute in Baltimore, told the Space Studies Board he was concerned about the technical readiness of any robotic servicing option, and preferred an astronaut mission. Dennis Wingo of SkyCorp believes this may no longer be an option, however. "The administrator has dug in his heels. So what we have is a compromise position between dropping Hubble in the ocean and risking another seven astronauts."

Weiler estimated that building the service robot would cost less than \$300 million, and said that even with a rocket launch and other technology development thrown in, robotic servicing would still be cheaper than a shuttle mission. That may make it the perfect way out of O'Keefe's dilemma — it might be riskier for the telescope, but not for the astronauts.

Link from hygiene to allergies gains support

Erika Check, Washington

Researchers have confirmed that a pristine environment can contribute to allergies in small children — and thrown up new questions about the underlying causes of allergic disease.

The 'hygiene hypothesis' proposes that germ-free environments lie behind the huge increase in asthma and allergies seen in the West since 1960. Studies presented this week at the annual meeting of the American Academy of Allergy, Asthma and Immunology in San Francisco fired debate on one aspect of the hypothesis: can immune responses to single allergens protect against other allergens in later life?

Researchers have found that keeping a cat in a young child's house reduces the risk of cat allergies in later life. But they disagree about whether cats also protect against non-cat allergies.

In a study of 224 children in New Zealand, those with a cat in the house reacted less to cat and dog allergens than they did to dust mite allergens, the meeting heard. The study group, led by Julian Crane of the Wellington School of Medicine and Health Sciences, found that different immune-system molecules responded to the cat allergens than to other allergens. Crane and his colleagues say this hints at the immune-system response that protects against cat allergies.

They also believe their work shows that the cat response is not universally protective. "There's a very strong cat effect that can be cat-specific," says immunologist Thomas Platts-Mills of the University of Virginia in Charlottesville, who worked on the study.

The data contradict a study in which children who lived in a home with two or more cats or dogs appeared to gain protection

against many allergies (D. R. Ownby *et al. J. Am. Med. Assoc.* **288**, 963–972; 2002). But despite the results of that study, many researchers argue that different allergens affect the immune system in distinct ways. For instance, research published this month showed that mice infected with the influenza virus were predisposed towards developing allergic asthma (M. E. Dahl *et al. Nature Immunol.* **5**, 337–343; 2004). In contrast, hepatitis A infections are associated with lower risks of allergies (J. J. McIntire *et al. Nature* **425**, 576; 2003). Many researchers are trying to figure out why different antigens give these conflicting results.

The answer may lie in how children are exposed to allergens, Ursula Krämer of the Environmental Health Research Institute in Dusseldorf, Germany, told the meeting. Krämer's team studied 3,241 German chil-

dren and found that those with more educated parents were more likely to develop cat allergies after keeping a cat in the house. The team thinks that the relationship might be due to behavioural differences, such as how much contact the child has with the pet.

Lacking a good understanding of the mechanisms behind allergic disease, scientists struggle to explain these studies. But Kenneth Adams, chief of the asthma and inflammation section at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, says that a more hygienic Western world has clearly influenced rates of allergic disease — but the effect also varies with different allergens, socioeconomic factors and genetics.

"The basic notion that infection can protect against sensitization is probably a huge oversimplification," says Adams. ■



Furry friends: pet cats can protect children against allergies, but it's not clear how broad the benefit is.

CHUCK SAVAGE/CORBIS

Biology hogs the science budget, senator complains

Erika Check, Washington

Money tensions on Capitol Hill erupted into farce last week when researchers at the National Institutes of Health (NIH) were branded "pigs" by a senior senator during a budget debate.

"The NIH is one of the best agencies in the world," Senator Pete Domenici (Republican, New Mexico) told colleagues at the 11 March debate. "But they have turned into pigs. You know, pigs! They cannot keep their oinks closed. They send a senator down there to argue as if they are broke." Observers said that Domenici also used his hand to mime a pig's snout in front of his face and wiggled his fingers. "Will you listen to what has happened to the NIH in

five years and tell me that they should get this much money?" he said.

Domenici was responding to Arlen Specter (Republican, Pennsylvania) — one of the NIH's main champions in the Senate — who successfully proposed that the budget resolution for 2005 incorporate a \$1.3 billion boost for the NIH. This is \$536 million more than President Bush has proposed for the agency. The resolution guides the appropriations subcommittees who determine actual spending levels.

But the proposal outraged Domenici, a strong supporter of physical-sciences research whose home state houses two huge nuclear weapons laboratories, Los Alamos and Sandia. Backers of the physical sciences

have become increasingly frustrated in recent years by the failure of other research agencies to attract the kind of increase obtained by the NIH. These feelings had seldom been publicly expressed, however — until Domenici's outburst.

The NIH is one of the few agencies apart from defence and homeland security that could get a big budget increase next year. On 2 February, President Bush proposed that the agency should get \$28.6 billion in the 2005 fiscal year, which starts in October, a 2.6% increase on the 2004 budget. But NIH advocates want 8–10%, to help the biomedical research agency sustain the momentum created by the doubling of its budget between 1998 and 2003. ■



The US delegation at December's meeting, where stalemate set in over where to build ITER.

Partners fail to find common ground for fusion project

Declan Butler, Paris

Iter means 'the way' in Latin — but 'lost its way' might be a more apt description of the international fusion project that bears the name.

A technical assessment intended to break a stalemate over where to site the facility has itself ended in deadlock, with the six partners blaming each other for the lack of progress.

Negotiations over ITER's site have been stalled since a ministerial meeting in Washington last December, when the United States and South Korea backed the Japanese site at Rokkasho, and China and Russia supported the European Union's site at Cadarache in France. The US\$5-billion project seeks to show that heating plasma in a magnetic field can produce fusion energy.

But a meeting of experts in Vienna, Austria, on 12 and 13 March ended in discord. Supporters of the European site insisted that the meeting should publish a technical comparison of the two sites — which they claim would have shown up their strengths.

Supporters of the Japanese site, including the United States, countered that the technical discussions were meant to feed into a wider political decision on a site.

The outcome of the meeting — privately described as 'fiery' by participants — was a bland statement that it had agreed to disagree. No further meetings have been scheduled to decide on a site, and plans to consider making the experimental reactor part of a broader, international fusion-research package remain just plans.

One Japanese government official put a brave face on the situation. "We're going up

the stairs one step at a time," he says. "There's no such thing as an endless staircase, so eventually we'll get there." A European fusion scientist close to the talks disagreed: "We're not going anywhere."

US sources familiar with the negotiations say that the government there wants to take a backseat, in the hope that Europe and Japan can come to an agreement. But many researchers wonder if an accord can be reached. "People in the United States have no idea how this will get resolved," says Stephen Dean, head of Fusion Power Associates, a Washington-based advocacy group for fusion research.

Others worry that a decision on location will not be reached by the summer, when US researchers are set to re-evaluate the nation's role in the project. "Europe and Japan have to arrive at a decision, otherwise this could go on forever," says Dale Meade, a physicist at the Princeton Plasma Physics Laboratory in New Jersey.

An eight-person European Union delegation went to Tokyo this week to seek agreement with Japanese experts and government officials. One European scientist privately branded the trip "mission impossible".

But despite the deadlock, India last week became the latest nation to say it might join the project. According to science secretary Valangiman Ramamurthy, it accepted an offer from David King, science adviser to Prime Minister Tony Blair, to join ITER as a partner of Britain, which will cost India less than joining the project as a full partner. ■

Additional reporting by Geoff Brumfiel, David Cyranoski and K. S. Jayaraman.

Ice machine sheds light on climate history written in dust

Jim Giles, Montreal

Astrophysicists working on an Antarctic particle-physics experiment have serendipitously created a device that can track ancient climate trends.

The researchers say that initial results from the probe, to be published in the next few weeks, could provide the most accurate record yet of global temperature change over the past million years. Climatologists think that average global temperature is closely correlated with levels of volcanic dust in the atmosphere — something they can track by careful examination of the polar ice caps.

The 'Dust Logger' arose from work in the 1980s by researchers interested in using polar ice in detecting subatomic particles called neutrinos. These stream in from space and occasionally collide with protons or neutrons in Earth, creating a shower of secondary particles. Researchers can identify the path of neutrinos by tracking light emitted when the new particles are formed.

The Antarctic Muon and Neutrino Detector Array (AMANDA) used this technique to spot its first neutrino in 2001. But to calibrate the detector, Buford Price, an astrophysicist at the University of California, Berkeley, had to read up on how the movement of light through ice is affected by dust trapped within it.

Price, together with his graduate student Ryan Bay, soon realized that they could generate data on dust levels in the ancient atmosphere by measuring the optical properties of ice below the Antarctic surface.

The first tests of the Dust Logger took place in 2001, using holes drilled to extract cores of ice for other palaeoclimate experiments. The logger sends out bursts of light that last a nanosecond, and by recording how long it takes for the light to be reflected back to the logger, Price and Bay can estimate the dust content at different depths.

The results were discussed at the March meeting of the American Physical Society in Montreal, Canada, and are due to appear in the *Proceedings of the National Academy of Sciences*. They track levels in an Antarctic bore hole 1 kilometre deep, which is the equivalent to looking back on about a million years of climate history. The logger revealed 60–70 peaks in dust levels thought to be caused by volcanic eruptions. Less than 20 such peaks had been identified from direct examination of the ice core, says Price. ■

Societies take united stand on journal access

Jim Giles, London

More than 40 biomedical societies have banded together to counter calls for them to provide immediate and unrestricted access to the scientific literature they publish.

On 16 March, the societies launched the Washington DC Principles for Free Access to Science, a set of guidelines that backs free access to papers but reserves the right of publishers to charge for them for several months after publication.

In London, meanwhile, publishing experts told a parliamentary select committee that scientific societies might have to cut back on their non-publishing activities if open access catches on.

Signatories to the Washington DC principles — most of them US-based societies — say they need publishing revenues to pay for everything else they do, from running conferences to educating the public. “Any profit that we make goes back into the development of the next generation of scientists,” says Martin Frank, executive director of the American Physiological Society and spokesman for the group.

The societies say they are responding to the launch of the Public Library of Science’s first journal, *PLoS Biology*, last October. The free online journal generates income by charging authors \$1,500 to publish a paper.



Free for all: the UK parliament joins the debate on payment for papers.

But the societies say that the backers of PLoS failed to acknowledge that both traditional and open-access publishing can coexist. “People were saying that PLoS was plotting the overthrow of the scientific publishing system,” says Frank.

Many society editors say it would be impossible to generate income for other activities if they only raised money through payments from authors. An international survey of 4,000 researchers’ attitudes to open-access publishing, released on 18 March by information scientists at City University, London, suggested that fewer than 5% of authors would be willing to pay more than \$1,000 to publish a paper.

Rhonda Oliver, managing director of Portland Press, the London-based publish-

ing arm of the Biochemical Society, says she would need to charge almost twice that just to keep the society’s journals afloat. Oliver says the publications also give a surplus of £500,000 (US\$920,000) to the society every year, and that the society would have to charge £3,000 for each paper published to keep that going.

Similar concerns have been raised by witnesses to the inquiry into scientific publishing run by the UK House of Commons’ Science and Technology Committee, which began hearings on 1 March and plans to report this summer.

But the editorial director of BioMed Central, a London-based publisher of more than 100 open-access journals disagrees: Peter Newmark argues that societies are basing estimates for author payments on existing publication systems. He says that the figure would be much lower if journals moved to online-only publication and factored in savings generated by not having to print journals or sell subscriptions.

Newmark also questions whether the societies should be channelling publication revenue into other activities. He says that grants and education programmes are valuable, but points out that the money ultimately comes from public sources. “They’re taxing the universities to support the societies,” he says. ■

US Army backs Swedish cell study

Alison Abbott, Munich

The US Department of Defense is giving \$240,000 to Lund University, Sweden, in order to fund research into treating Parkinson’s disease with human embryonic stem cells.

The unlikely partnership came about after Patrik Brundin, the Swedish project’s principal investigator, was advised to submit a proposal to the US Army’s Neurotoxin Exposure Treatment Research Program. The recommendation came from the New York-based Michael J. Fox Foundation, which supports research into Parkinson’s.

The army’s programme supports research into neurological diseases in case neurotoxins are used as weapons of war. It considers grant applications from researchers anywhere in the world, as do many US government research programmes.

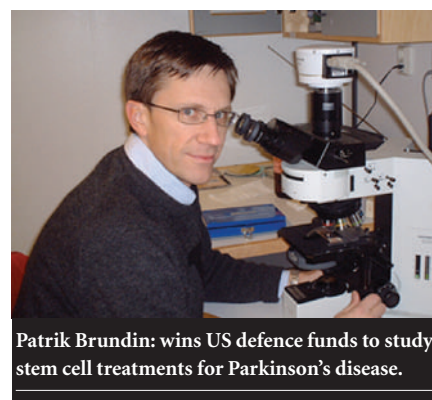
Parkinson’s disease occurs when dopamine-producing cells degenerate in a brain area called the substantia nigra.

Brundin was involved in pioneering studies in the late 1980s that tried to replace those cells using substantia nigra tissue from aborted human embryos.

In the small series of clinical trials in Sweden and the United States, some patients did very well, some less well, some did not respond and a small number suffered the side effect called dyskinesia, where a patient has uncontrolled movement of the head or limbs.

“We do not understand the basis of this inherent variability,” says Brundin, who points out that the supply of embryonic tissue is unreliable and could be one source of variation. Human embryonic stem cells that are stimulated to differentiate into dopamine-producing cells could provide more reliable material for transplantation.

His team will first develop methods of differentiating a stem cell line produced at the University of Gothenburg, Sweden. The cell line is one of those approved for federal



Patrik Brundin: wins US defence funds to study stem cell treatments for Parkinson’s disease.

funding by the US National Institutes of Health. These cells will be transplanted into the brains of rats which have a type of parkinsonism induced by chemical damage. The condition of the rats will be continuously monitored, as will the survival and correct biological functioning of the transplanted cells.

“In this way we hope to get some insight, under very controlled research conditions, into the variability of transplanted tissue in parkinsonism,” Brundin says. ■



Kiyoshi Kurokawa (standing) wants Japan to implement its own version of the US National Academies.

Japan shakes up council to offer scientists political clout

David Cyranoski, Tokyo

Don't hold your breath — but Japanese scientists may be on the verge of getting a strong voice in their nation's policy-making.

At a parliamentary committee hearing on 19 March, the country's minister of science and technology and the president of the Science Council of Japan (SCJ) proposed moving the currently ineffectual council — which represents some 760,000 scientists from 1,481 academic societies — from the ministry of home affairs to the prime minister's office.

SCJ president Kiyoshi Kurokawa claims that the transfer could put the council on a par with the US National Academies, which advise the government on many topics related to science, medicine and engineering.

This would give the council the status it was intended to have, advocates of the change say. Since its foundation in 1949, the SCJ has reported to the government on issues from mad cow disease to scientific misconduct. But with some exceptions, its advice has been ignored, says Kurokawa. Ministries instead took advice from their scientific committees, he says.

"The SCJ has always been an advisory body — now it would be a real one," the SCJ's executive director, Masatsugu Yoshida, commented at the hearing.

Kurokawa says that central government needs a voice that is not compromised by ministerial affiliation. The existing advisory committees are "a protective mechanism for bureaucrats when policies fail", he says.

But whether the promised influence and independence will materialize remains to be seen. Yukio Hatoyama, a prominent representative of the opposition Democratic Party, agrees that the council should be reformed.

"The purpose of the council isn't clear to the people," he told the hearing. "It has a meagre existence." Hatoyama remains unconvinced that the proposed reforms will enable the council to give independent advice on sensitive topics, as it will be attached to the prime minister's office and reliant on it for funding and administrative support.

Toshimitsu Motegi, the science minister, argued that the reformed council would be independent. Kurokawa added that the body would be transparent in all of its functions and would post its reports on the Internet.

Under the proposal, the SCJ would work with the Council for Science and Technology Policy (CSTP), the country's top science policy-making body. But the CSTP has been criticized as being out of touch with science. Its 15 members include the prime minister, six ministers of state, and eight private citizens appointed by the government, few of whom are working researchers.

The proposal also seeks to allow the SCJ's 210 members to elect new members directly. The current method — taking people recommended by member societies — creates an environment where members are seen as representing their society's interests, says Kurokawa. The council's seven existing divisions would be combined into three: human and social sciences; life sciences; and physical sciences and engineering.

The diet's lower house will meet again this week and is expected to approve the proposal and pass it on to the upper house. It is likely to become law in April. In preparation for the SCJ's beefed-up role, Kurokawa is setting up a joint committee with the US National Academies to look at topics related to national security.

Sellafield seeks solid foundation for cleaner wastewater

Laura Nelson, London

Persistent discharges of radioactive waste into the Irish Sea could come to an end thanks to a technique that can lift an awkward pollutant out of water.

Chemists at the Sellafield reprocessing plant in northwest England have tested a process aimed at removing technetium-99 from wastewater.

Technetium-99 has a half-life of 213,000 years and is the main radionuclide still being discharged from Sellafield. It has been accumulating in seaweed and lobsters as far afield as the North Sea, leading to angry protests from the Norwegian and Irish governments.

Most of Sellafield's technetium-99 was generated by a now-disused processing plant, and some 1,800 m³ of contaminated wastewater is stored in tanks at the site. The isotope had proved awkward to extract and so the water, once cleaned of other pollutants, is being dumped into the sea at a rate of about 240 m³ per year.

The new cleaning process uses the salt tetraphenylphosphonium bromide (TPP), which is added to the wastewater. TPP binds with technetium-99 to form an insoluble compound that precipitates out as a solid.

The Sellafield chemists completed tests on the technique in December and will release their evaluation in April. If the results are positive — as many observers expect — BNFL, which operates the plant, is set to implement the technology at a cost of £2 million (US\$3.7 million).

"The results are very promising but we still have to go through a final assessment," says Ian Parker, head of nuclear regulation at the UK Environment Agency, which oversees the plant.

Parker adds that the agency had been worried that the precipitate might have been difficult to incorporate in concrete for safe and permanent disposal, but that early test results suggest this can be done.

Technetium-99 is not thought to be harmful in small doses, but Ireland and Norway have complained that traces of the isotope found in marine life have led to cuts in the prices fish merchants will pay for lobsters and other seafood.

The abatement technology could ease a major headache for BNFL, but other problems remain at Sellafield. European courts, for example, have been hearing a plea from the Irish government for the facility to close its mixed oxide fuel plant, which releases small amounts of plutonium into the sea.



Listening in: the Allen Telescope Array will be ready to seek extraterrestrial signals by 2010.

Microsoft co-founder dishes out millions to search for aliens

San Diego The Search for Extraterrestrial Intelligence (SETI) has been boosted by a US\$13.5-million donation from Microsoft co-founder Paul Allen.

Some of Allen's cash will pay for 32 radio dishes, with the remainder pledged to a further 174 dishes, as long as SETI can raise \$16 million from other sources. If all goes well, an array of 350 six-metre dishes will be completed by the end of the decade.

The Allen Telescope Array, to be built at Hat Creek Radio Observatory, 300 miles northeast of San Francisco, would join the

SETI Institute in Mountain View, California, in scanning distant galaxies for signs of intelligence. It will also be used for other radio-astronomy research. Allen paid \$11.5 million for a three-year planning project to design the array, a joint venture between the SETI Institute and the Radio Astronomy Laboratory of the University of California, Berkeley.

Centralized network set up to run UK clinical trials

London In a bid to strengthen the biotechnology and pharmaceutical industries, the UK government announced in its budget on 17 March the formation of a National Clinical Research Network to coordinate clinical trials. The network, set up to centralize trials' funding and regulation, follows recommendations made last November by a government-backed report from the Bioscience Innovation and Growth Team, a group of researchers, industry representatives and government officials (see *Nature* 426, 216; 2003).

The changes will be part of a ten-year framework for medical science that was announced by Chancellor Gordon Brown on the day before the budget. The clinical research network will bring together pharmaceutical companies, charities, the

government and the National Health Service.

The government also pledged to spend an extra £100 million (US\$180 million) on research and development in the National Health Service by 2008, taking the service's total research budget to £1.2 billion.

Bird-flu researchers warned of complacency

Paris Asian countries should be wary of declaring victory over avian flu, health officials warned last week. They fear that complacency will make a human pandemic more likely.

The warning followed signs in Vietnam, Thailand and China that culling and vaccination is bringing the bird-flu outbreak under control. But in a joint statement issued on 19 March, the Food and Agriculture Organization of the United Nations and the World Organisation for Animal Health said that as long as the highly pathogenic H5N1 bird flu virus is circulating in Asia, the risk of a human pandemic remains high, and that countries should prove that they are virus-free before restocking poultry.

Experts at a World Health Organization meeting in Geneva, Switzerland, last week also highlighted the lack of preparation for a human flu pandemic. Delegates emphasized the need for better surveillance of animal



Don't count your chickens: the H5N1 bird flu virus could still cause a human pandemic.

and human influenza strains, and the creation of antiviral drug stockpiles and systems to accelerate vaccine production.

Arizona observatory goes ahead on alternative site

Washington Construction has begun on a beleaguered observatory project, which has been moved from its original site after opposition from a Native American group.

The VERITAS (Very Energetic Radiation Imaging Telescope Array System) observatory, whose four 12-metre telescopes will be among the most sensitive in the world for detecting high-energy γ -rays, was originally to reside at the foot of Mount Hopkins in southern Arizona. But several

tribes of Native Americans who operate a nearby ceremonial 'sweat lodge' objected to sharing the mountain with astronomers.

Local US Forest Service officials ruled that VERITAS would not interfere with the lodge, but the ruling was overturned last April by the service's southwest regional office after Native American and environmental groups appealed. The Smithsonian Astrophysical Observatory, lead partner in the international VERITAS consortium, opted to shift the \$13-million facility to Kitt Peak outside Tucson, Arizona. With US funding from the Department of Energy and the National Science Foundation, the project has already built a prototype telescope and camera. These will be the first elements in the array when it moves to Kitt Peak in 2006.

Genome microarray has bacteria on the slide

New York A microarray chip that holds up to 30,000 genomes on a single slide could help researchers decipher which bacterial genes cause disease, according to scientists at the University of Michigan in Ann Arbor.

Most gene chips are printed with an array of tiny dots, each made up of a separate gene from a single genome. On the prototype chip (L. Zhang *et al.* *BMC Microbiol.* 4, 12; 2004), each dot is instead

a mixture of all the genes in an entire bacterial genome. Bacterial strains of the same species can vary genetically by up to 25%, and these differences can determine virulence and the type of host they infect.

The prototype slide currently carries only spots from *Escherichia coli*, but the authors of the study plan to expand the selection to include other species, with the aim of assessing the function of uncharacterized genes.

Web debate: access to scientific literature

The Internet is changing how scientists work and publish. Publishers are testing new business models, including open access, in which the author pays and content is free to the user. This week, *Nature* launches an ongoing web focus to explore current trends and future possibilities. Each week, the website will publish specially commissioned comment and analysis from leading scientists, librarians, publishers and other stakeholders. We will include key links and articles from our archive.

This week's debate includes articles by co-founder of the Public Library of Science Patrick Brown and *Science*'s editor-in-chief Donald Kennedy. All content is available free.

► www.nature.com/nature/focus/accessdebate

Feast or famine?

Proponents call it a miracle. Detractors call it smoke and mirrors. Will the System of Rice Intensification feed the hungry third world or needlessly divert farmers from tried and true techniques? Christopher Surridge investigates.

Rice feeds more than half the people in the world; but not well and not for much longer. As the population rises, so does the demand for rice, yet yields of the crop are levelling out. Already, more than 400 million people endure chronic hunger in rice-producing areas of Asia, Africa and South America. And demand is expected to rise by a further 38% within 30 years, according to the United Nations. To call attention to the problem, the UN has declared 2004 the International Year of Rice. "Rice is on the front line in the fight against world hunger and poverty," says Jacques Diouf, director-general of the Food and Agriculture Organization of the UN.

But no amount of fanfare can alter the disheartening conclusion drawn by a growing number of agronomists — rice yields are approaching their limit. Despite efforts on a variety of fronts, including genetic engineering of rice strains for improved nutrition and growth¹, no solution has been found.

So it is no surprise that a simple method that claims to boost yields at lower cost to farmers is being hailed by many as that solution. The System of Rice Intensification (SRI), developed in the late 1980s in Madagascar, has since been adopted in countries from Sri Lanka to Sierra Leone. In Cambodia, SRI was unheard of in 2000, but by last year nearly 10,000 farms had converted to it, a figure that may reach 50,000 this year.

But although advocates of SRI routinely report yields up to twice those achieved by conventional agriculture, many eminent agronomists dismiss such achievements as the result of poor record keeping and unscientific thinking. A new set of field trials published this month² seems to support this view



Wonder rice? The System of Rice Intensification involves more sparsely planted, drier fields than is normal. Its supporters believe it gives more than double the yields of conventional farming.

unequivocally. Yet proponents are standing firm, and the dispute over how best to grow rice seems set to continue.

SRI was developed nearly 20 years ago by Father Henri de Laulanié, a Jesuit priest who worked with farming communities in Madagascar from 1961 until his death in 1995. In conventional rice agriculture, the plants spend most of the season partially submerged. During a 1983 drought, which prevented many farmers from flooding their paddies, de Laulanié noticed that the rice plants — particularly their roots — showed unusually vigorous growth.

Intensive care

From this and other observations, de Laulanié developed the three main tenets of SRI: that rice seedlings should be transplanted quickly when young; that rice plants should be spaced widely apart; and, most importantly, that the rice fields should be kept moist but not flooded. Flooding, he believed, stunted the plants' roots by preventing proper soil aeration. In addition to these principles, de Laulanié took a page from organic agriculture, emphasizing plant husbandry over the use of chemical fertilizers, so that poor and rich farmers alike could practise SRI. To help promote his method, in 1990 de Laulanié established a non-governmental organization called Association Tefy Saina, which means "to improve the mind".

Despite de Laulanié's successes, SRI would have remained a Madagascan curiosity but for Norman Uphoff, a political scien-

tist and director of the International Institute for Food, Agriculture and Development at Cornell University in Ithaca, New York. In 1993, Uphoff was part of a team attempting to find alternatives to the slash and burn agriculture that was destroying Madagascar's rain forests. It was clear to Uphoff that finding a way to increase rice yields in the area, then about 2 tonnes per hectare, would greatly slow the clearing of the forests.

Then he met representatives of Association Tefy Saina. "I was hoping for a yield of 4 tonnes per hectare, so when they said they could get 15 or more, I frankly didn't believe them," Uphoff says. Such doubts evaporated once farmers in the rainforest regions started using SRI. The results, says Uphoff, were stunning. "By the end of the second growing season we were getting over 8 tonnes per hectare." In 1997, Uphoff began promoting SRI throughout Asia. It is largely through his efforts that the practice has become so widespread.

But some observers say SRI's proponents do not have the evidence to back up their claims. John Sheehy, an agronomist at the International Rice Research Institute (IRRI) in Manila, the Philippines, points out that most SRI field studies have appeared in conference proceedings and other publications not subject to peer review. Although there are exceptions³, such reports, Sheehy says, are little more than anecdotes without details of data collection methods, nitrogen inputs or other measures that a critical reader would need to judge the success of a trial.

Peer review is particularly important for



Against the grain: Henri de Laulanié (above) invented the System of Rice Intensification. John Sheehy (left) has found it wanting.



field studies, because of the pitfalls involved in assessing yields, says Achim Dobermann, a soil scientist at the University of Nebraska, Lincoln, who has attempted a critical assessment of SRI⁴. One concern is the rice's water content. Under SRI, rice takes about two weeks longer to mature for harvest than rice grown in conventional systems, by which time the grain has taken up much more water. Unless the grain is carefully dried, the SRI field will seem to have yielded more rice, when in fact the increase could be water.

This month, Sheehy, Dobermann and IRRI researcher Shaobing Peng have published their own trials of SRI². At three experimental stations in Hunan, Guangdong and Jiangsu provinces of China, they grew rice using SRI and conventional best practice on

living-room-sized plots in the same fields. Weeds were suppressed with herbicides on the conventional plots but pulled by hand in SRI plots. SRI plots received extra rapeseed cake fertilizer. Conventional plots were flooded; SRI plots were not. Yet overall, no significant differences were found between the two cropping systems. SRI yielded slightly higher in Jiangsu, but fared slightly worse in Hunan.

On trial

Dobermann is not surprised. Every component of SRI has been studied before and found to have little effect, he says. The results also fit Sheehy's theoretical calculation of how much rice a field can produce. He believes the upper limit is set by the amount of sunlight that falls on it. Based on weather data for Madagascar, Sheehy has calculated theoretical maximum outputs for areas that have reported the most impressive yields under SRI. By his estimates, the reported yields are as much as 10 tonnes more than is possible². "You can't get out more than gets put in," he says.

So what accounts for SRI's growing popularity? Dobermann suspects practitioners have been lulled by the apparently greater vigour of their plants. When rice plants are grown close together, as in conventional practice, they produce fewer tillers, the side branches upon which the grains are formed. With the wider spacing of SRI, the individual plants produce more tillers and so more grain. But this merely compensates for hav-

ing fewer plants per hectare, and the yield is not significantly affected.

Others suspect that SRI may perform better in some regions than others. John Angus, a crop scientist with the CSIRO, Australia's national research agency in Canberra, thinks SRI could increase yields only under specific conditions. The soils in the highlands of Madagascar are highly reducing — they are acidic and low in oxygen. Reducing conditions also occur in parts of Japan and the Australian rice belt where Angus says it has proved helpful to drain paddies in the middle of the growing season to allow more oxygen into the soil.

Proponents contend that SRI owes its popularity to impressive yields. T. M. Thiagarajan, dean of the Agricultural College and Research Institute in Killikulam, India, rejects criticisms of individual aspects of SRI. In combination, he says, the whole is greater than the sum of its parts: "The synergistic effect of all these components is the crucial thing." Thiagarajan helped convince the Tamil Nadu state government to spend US\$50,000 on promoting SRI to local farmers.

Roots of farming

Uphoff maintains that critics' assumptions are too firmly rooted in conventional practice. Models for estimating maximum yields will not necessarily translate to SRI, for instance. "The coefficients for the calculations are based on plants with stunted root systems. SRI plants have extensive root systems," he says.

Nor will single-season field trials reveal the full potential of SRI, he argues, because over time, better oxygenation leads to the build-up of soil bacteria that interact with the roots and improve the condition of the soil. Even if SRI fails to increase yields when first introduced — the case in Thailand, for instance — further seasons will see it come into its own.

In some ways, the debate resembles that currently raging over organic agriculture. For advocates, SRI is a grassroots movement to resist the influence of global agribusiness by reducing dependence on chemical inputs. Detractors call it a waste of time that is diverting resources from more promising approaches such as genetic engineering.

One thing both sides agree on is that meeting the UN's goal of halving hunger and poverty by 2015 will require radical changes in agriculture. Whether this will come from the uplands of Madagascar or the modern agronomy laboratory remains an open question.

Christopher Surridge is a senior biology editor at Nature.

1. Surridge, C. *Nature* 416, 576–578 (2002).
2. Sheehy, J. E. *et al.* *Field Crops Res.* doi:10.1016/j.fcr.2003.12.006 (2004).
3. Stoop, W. A., Uphoff, N. & Kassam, A. *Agric. Syst.* 71, 249–274 (2002).
4. Dobermann, A. *Agric. Syst.* 79, 261–281 (2004).



The sweet smell of success

Smell is arguably the most evocative and mysterious of our senses. But thanks to advances in our understanding of the cells that detect odour, its secrets should now start to be revealed. Carina Dennis sniffs around.

Lawrence Katz likes to start his lectures to medical students by opening a vial containing the pig version of the pheromone androstenone, used by boars to attract their mates. Katz, a neuroscientist at Duke University Medical Center in Durham, North Carolina, can't smell it himself. But many of his students retch in their seats at the pungent odour. "My graduate student says it's like having his head shoved in a urinal," says Katz.

Such individual differences in sensitivity to smells are thought to be the result of variations between olfactory receptors — proteins carried on the surface of sensory neurons that detect volatile chemicals wafting up our nasal passages. We have hundreds of distinct receptors, which between them can distinguish thousands of different chemicals. But without a clear understanding of how this information is encoded by olfactory receptors, scientists have made little progress in determining how our brain perceives complex fragrances such as 'chocolate' or 'freshly baked bread'.

"The mammalian nose is the best chemical detector in the world, yet we still don't under-

stand how it codes," says Stuart Firestein, a neuroscientist at Columbia University in New York, whose team was the first to document the molecular interaction between a specific receptor and a particular odorant¹.

That should now start to change. Experiments are beginning to reveal how individual sensory cells end up carrying just one type of receptor. Researchers should also soon be able to culture cells featuring any receptor they want. This will make it easier to work out which receptors respond to which odours. And that, in turn, should help scientists to track the neural processing of smell, all the way from sensory cells to the brain's cortex.

Odour eaters

Our sense of smell begins with two tiny patches of tissue that line the roof of our nasal passages, just below eye level (see Figure, opposite). Crammed into these patches, which together are about as big as a postage stamp, are the ends of some 10 million sensory neurons. Minuscule hair-like extensions, called cilia, project from these cells and carry the olfactory receptors.

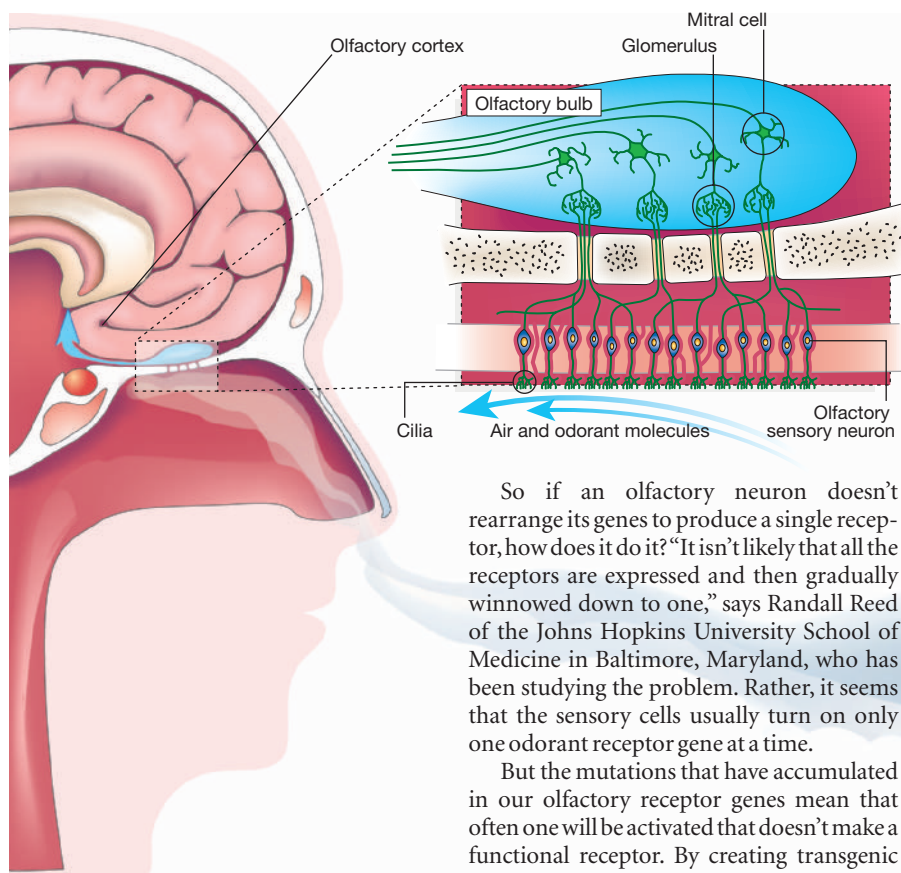
These receptors are encoded by the

largest family of genes known to exist in mammals. Mice, which rely heavily on their sense of smell, have about 1,200 of these genes, most of which are in full working order². By contrast, nearly two-thirds of the roughly 1,000 human olfactory receptor genes, which are spread over all but two of our chromosomes, have accumulated mutations that render them useless³. It seems that we have been losing functional olfactory receptors ever since our primate ancestors evolved full colour vision⁴.

But even this diminished sense of smell is fiendishly difficult to investigate compared with vision. It is relatively easy to determine what an individual sensory cell in the retina responds to, in terms of wavelengths of light and a defined area within the field of vision. Over the past few decades, neuroscientists have used electrophysiological recordings to work out how responses are combined and processed to perceive more complex visual patterns.

Olfactory scientists have not been able to build from such firm foundations. A single odorant chemical can be recognized by multiple olfactory receptors, and each recep-

Follow your nose: how we detect odours



So if an olfactory neuron doesn't rearrange its genes to produce a single receptor, how does it do it? "It isn't likely that all the receptors are expressed and then gradually winnowed down to one," says Randall Reed of the Johns Hopkins University School of Medicine in Baltimore, Maryland, who has been studying the problem. Rather, it seems that the sensory cells usually turn on only one odorant receptor gene at a time.

But the mutations that have accumulated in our olfactory receptor genes mean that often one will be activated that doesn't make a functional receptor. By creating transgenic mice bearing genes that encode either a working or a non-functional receptor, teams led by Reed and by Hitoshi Sakano at the University of Tokyo have independently shown that once a sensory neuron hits upon a gene that encodes a working receptor, no others are activated^{8,9}. The cell's fate then seems to become fixed, and it carries on producing the same receptor for the rest of its life⁷.

But how the cells ensure that multiple genes are not turned on at once remains unclear. Perhaps the genes must interact at random with a molecule that is only available in very limited quantities, such as a factor needed to free DNA from its dense packing within a chromosome. Sakano has another

theory. He suspects that olfactory receptor genes need to interact with a stretch of DNA, known as a locus control region (LCR), located some distance away on the same chromosome. This must loop around and snuggle up against the olfactory genes, Sakano argues, but can only do so with one gene at a time. For one cluster of olfactory receptor genes, he has already identified a candidate LCR, which he calls the 'H' region. When it is deleted, the genes in the cluster remain inactive⁸.

Knowing how individual sensory cells come to produce one particular receptor will still leave smell researchers in the dark about how the cells collectively encode odours. To help crack this code, scientists want to study the interaction between receptors and different odourants for individual cells. But olfactory sensory cells are hard to culture and handle in the lab. "They are difficult to grow and nearly impossible to separate from their neighbours," says Mombaerts. That poses a big problem for researchers trying to crack the olfactory code, as neighbouring cells bear different receptors.

Follow the scent

An alternative would be to genetically engineer cells that are more amenable to culturing in the lab to make them carry particular olfactory receptors on their surfaces. But attempts to do this have been frustrated by the failure of the olfactory receptors to transfer to the outer membrane of the engineered cells.

Hiroaki Matsunami at Duke University may now have removed this experimental bottleneck. In unpublished work, he has figured out that olfactory receptors need a little help from a specific protein to find their way to the cell's surface. If the receptor genes are introduced into cells together with the gene encoding this protein, they can find their way to the exterior. Matsunami is beginning to use these 'hana' cells — the name derives from the Japanese for nose — to identify which odourant molecules are recognized by different receptors.

That won't be a simple task. New evidence from a team led by Kazushige Touhara at the University of Tokyo's Graduate School of Frontier Sciences in Chiba suggests that some odours can also inhibit olfactory receptors, rather than activate them¹⁰. This means that our nasal cavities represent a competitive battleground where odours jockey with each other to stimulate or block olfactory receptors. This concept has been recognized by perfume companies, which add molecules to their products to enhance a fragrance.

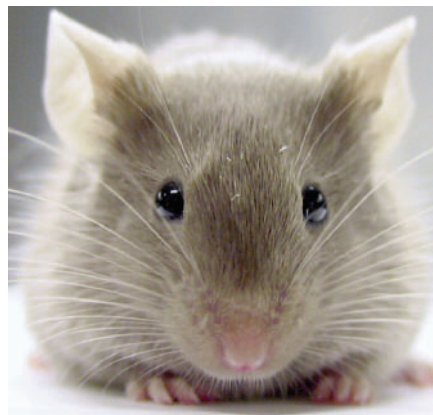
As more odours are added to a mixture, some will cancel others out or change how we perceive the mixture's smell. "We have a physiological limit of about three in the number of odours we can distinguish at once," says David Laing, a neuroscientist at

tor is thought to recognize several odorants. Individual cells carry only one type of receptor, which means that the detection of a particular odorant is encoded by the firing of a distinct combination of sensory cells. But that code remains unbroken: smell researchers have identified the volatile odorant partners of only about a dozen mammalian olfactory receptors.

One piece at a time

More fundamentally, researchers haven't been able to work out how each sensory neuron comes to bear just one type of receptor. One idea was that the mechanism might involve an irreversible rearrangement of the DNA of olfactory receptor genes, in much the same way as the B cells of our immune system cut and paste their genes to produce just one particular antibody.

But there was little supporting evidence, and cloning technology has now disproved this theory. If you clone a mouse from a single B cell, its entire immune system can generate only one antibody⁵. Two groups have repeated the same experiment with mature olfactory neurons, and shown that the resulting mice have the normal diversity of olfactory receptors^{6,7}. "Cloning was a straightforward way to test the question," says Peter Mombaerts of Rockefeller University in New York, whose team reports its results in this issue of *Nature*⁷.



Smelly beast: this mouse, cloned from an olfactory neuron, has a normal sense of smell.

the University of New South Wales in Sydney, Australia¹¹. For instance, if you were given a vial with equal concentrations of banana, vanilla and strawberry smells, you will be able to distinguish all three odours. But throw in peppermint and chances are you won't be able to discriminate between all four. And the combination of odours may smell completely different from any of the individual components. "We can't predict what we are going to smell with various combinations of odours," says Laing.

But if smell researchers do crack the olfactory receptor code, neuroscientists interested in the sense of smell will find themselves in the position that their colleagues who study vision have enjoyed for several decades. They will be able to study various cells within the system, and relate the coding of information at the level of the receptors to higher levels of processing.

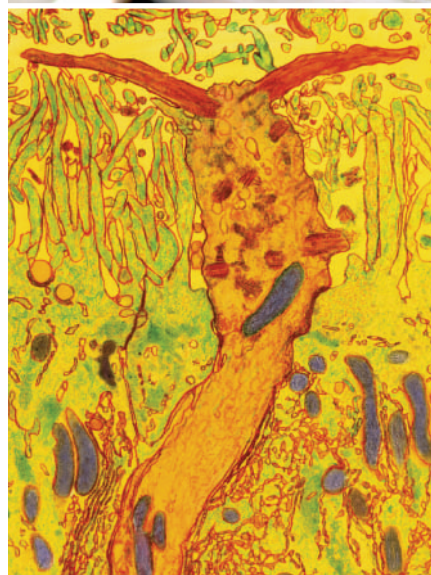
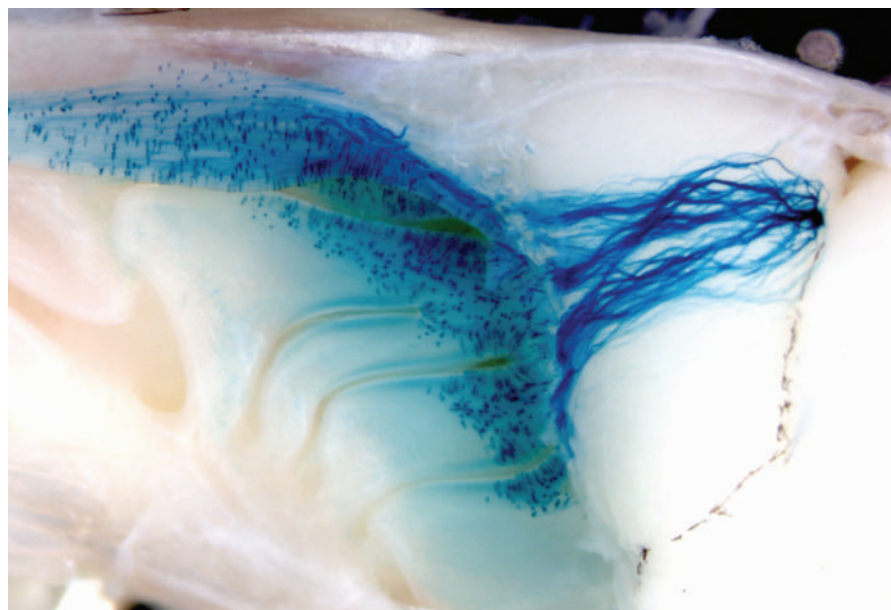
Smells like team spirit

The basic wiring of the olfactory system is already known. Sensory neurons send projections back to the olfactory bulbs, a pair of pea-sized structures nestled on the underside of the frontal lobe of the brain, set back behind our eyes. Each bulb is divided into two halves, and each of these halves contains hundreds of structures called glomeruli. These act like neural junction boxes linking the projections from sensory cells to neurons called mitral cells that relay information about smell to olfactory centres in the brain.

Sensory cells bearing the same receptor all converge in the same glomeruli, and each half of each olfactory bulb contains at least one glomerulus for each type of receptor. As a result, an odour that stimulates a specific group of sensory neurons will also cause particular glomeruli to fire — odours can be thought of as being coded through a 'spatial map' involving the firing of glomeruli at specific positions in the olfactory bulbs.

From here, information is relayed back to olfactory centres of the brain's cortex, which interpret the odours and direct our responses to them. But opinions differ on the extent to which this processing begins in the olfactory bulbs themselves.

Two research groups conducted different experiments to measure the input and output of the insect equivalent of the mammalian olfactory bulb, called the antennal lobe, and arrived at very different conclusions. From their physiological recordings, Richard Axel of Columbia University and his colleagues argue that the spatial map encoded by the antennal lobe is faithfully relayed by the neurons leaving the structure¹². But Gilles Laurent of the California Institute of Technology in Pasadena has evidence that the output of the antennal lobe looks very different from the input, with the information spread across a much greater number of outgoing neurons^{13,14}. He believes that not only do the out-



Nasal passage: smells are picked up by receptors on cilia (left, dark orange) at the end of a sensory neuron (light orange). Sensory cells bearing the same receptor (stained blue, above) then connect to the same glomerulus (top right).

going neurons tell the brain which receptors were activated, but that they also provide information about an odour through the timing of the signals they transmit¹⁵.

Whatever the case, it is unclear whether what happens in insects also applies to mammals. "The insect antennal lobe is thought to be quite similar to the mammalian olfactory bulb, but that still remains to be shown," says Laurent.

Higher levels of processing in the brain's olfactory cortex remain, for now, obscure. Some researchers argue that the neural connections to the cortex are so diffuse that it is impossible to discern any spatial map of odour coding. But Linda Buck of the Fred Hutchinson Cancer Research Center in Seattle, Washington, has used genetic tricks to track the flow of information to the cortex in response to the activation of particular olfactory receptors. Her findings suggest that input from sensory neurons bearing a particular olfactory receptor is relayed to discrete clusters of neurons in the cortex¹⁶.

But the resulting spatial map is different from that in the olfactory bulbs. For a start, it seems that olfactory neurons bearing a particular receptor can cause activity in several distinct regions of the cortex. "There is a real transformation of information in the olfactory cortex — and it's very complex," says Buck.

That complexity is likely to involve emotional responses and information from our memories. Olfactory scientists still have much work to do before they understand exactly how the experience of a particular odour can have the dramatic effect of Katz's boar pheromone. But thanks to the advances that are now being made, it is at least starting to look like a tractable problem. ■

Carina Dennis is Nature's Australasian correspondent.

1. Zhao, H. *et al. Science* **279**, 237–242 (1998).
2. Zhang, X. & Firestein, S. *Nature Neurosci.* **5**, 124–133 (2002).
3. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. *Genome Res.* **11**, 685–702 (2001).
4. Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D. & Pääbo, S. *PLoS Biol.* **2**, 120–125 (2004).
5. Hochedlinger, K. & Jaenisch, R. *Nature* **415**, 1035–1038 (2002).
6. Eggan, K. *et al. Nature* **428**, 44–49 (2004).
7. Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. *Nature* **428**, 393–399 (2004).
8. Serizawa, S. *et al. Science* **302**, 2088–2094 (2003).
9. Lewcock, J. W. & Reed, R. R. *Proc. Natl Acad. Sci. USA* **101**, 1069–1074 (2004).
10. Oka, Y., Omura, M., Kataoka, H. & Touhara, K. *EMBO J.* **23**, 120–126 (2004).
11. Livermore, A. & Laing, D. G. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 267–277 (1996).
12. Wang, J. W., Wong, A. M., Flores, J., Vossahl, L. B. & Axel, R. *Cell* **112**, 271–282 (2003).
13. Perez-Orive, J. *et al. Science* **297**, 359–365 (2002).
14. Wilson, R. I., Turner, G. C. & Laurent, G. *Science* **303**, 366–370 (2004).
15. Stopfer, M., Jayaraman, V. & Laurent, G. *Neuron* **39**, 991–1004 (2003).
16. Zou, Z., Horowitz, L. F., Montmayeur, J.-P., Snapper, S. & Buck, L. B. *Nature* **414**, 173–179 (2001).

New law does little to ease research pain in Spain

Validation of foreign degrees is still mired in expensive, time-consuming bureaucracy.

Sir—We read with interest your editorial “Ending the pain in Spain” (*Nature* **428**, 1; 2004) but feel that it did not adequately address the bureaucratic difficulties facing foreign-educated scientists when applying for jobs in Spain.

Applicants for government-funded permanent positions must provide either a Spanish doctorate or the Spanish (“homologated”) equivalent of a foreign doctorate. This homologation procedure is expensive (several hundred euros for taxes, translation and other fees), involves excessive paperwork, and is time-consuming, taking six months in theory but many years in practice.

The Spanish Ministry of Education passed a law on 20 February to address this issue, but the law does not provide much relief. Careful reading reveals few significant changes from the previous law.

Diplomas from foreign universities may now be validated in general, rather than for each individual case. This is a small step in the right direction, but if Spain is serious about improving the mobility of researchers, it should consider following EU directive 89/48/CEE and scrap the homologation requirement entirely. Failing that, it should remove the requirement to homologate the undergraduate degree, which requires the same paperwork and processing times, and

consider accepting documents in English, as many European universities already do.

We have sent an open letter to the Spanish ministers of education and of science, asking them to consider removing these remaining barriers both to Spanish scientists who wish to return home and to any scientist with the required expertise and will to come and work in Spain.

Mark van Raaij*

Departamento de Bioquímica, Facultad de Farmacia, Universidad de Santiago, 15782 Santiago de Compostela, Spain

*Signed on behalf of the Asociación Nacional de Investigadores

Ramón y Cajal and 89 international co-authors, whose names and contact details are available directly from the author.

Spain: politicians need to challenge the status quo

Sir—As a Spanish postdoc working outside Spain for almost five years, I welcome your Editorial “Ending the pain in Spain” (*Nature* **428**, 1; 2004).

As you say, science rarely occupies the headlines in the Spanish press. When it does, the story is often a comment on the achievements of Spanish researchers working abroad. As things improve, more scientists will return to Spain, but not as quickly as you suggest.

For example, it is not true that “Over the past three years, [the Ramón y Cajal programme] has repatriated almost 2,000 of Spain’s diaspora of postdocs”. If you look more carefully at the numbers, you will see that most of the 1,944 positions offered went to Spanish scientists already living in Spain, whereas only 21.4% were awarded to Spanish researchers living abroad. To be fair, the programme was not aimed solely at repatriating Spanish researchers, but intended to increase the overall number of researchers working in Spain by providing steady five-year contracts to both Spanish and non-Spanish nationals.

There are several reasons why few Spaniards working abroad have been attracted to this scheme. One of these is that their own research initiatives will often be subordinated to that of the group that receives them. This is understandable, to some extent, because this group will pay part of their salaries and provide them with space and equipment. Another is the uncertain future of people holding such positions. Contrary to the suggestion in your Editorial, the competition for funds in Spain could hardly be stiffer. Numerous people have returned to Spain, attracted by

schemes operating before the Ramón y Cajal programme, and were then abandoned by the system. Many of these had to emigrate again. This situation will continue until our politicians show an interest in challenging the status quo.

Miguel Ortiz Lombardía

York Structural Biology Laboratory, University of York, Heslington, York YO10 5YW, UK

Fusion: Bush agrees it’s time to end the impasse

Sir—I am puzzled by your Editorial “Time for Japan to shine?” (*Nature* **427**, 763; 2004) on resolving the stalemate over the choice of a site for the fusion project ITER. You analyse the situation with the Japanese and European bids fairly and accurately. You conclude that if one or the other is technically superior, the facility should be sited there, but that in the event of a tie or a near tie, then Japan should be the chosen site.

I agree with your analysis and have been saying the same thing for some time.

My puzzlement comes from the last paragraph in which you say “Europe should turn the other cheek to the Bush administration’s mischief-making and break the impasse”. This position, although new to *Nature*, is one that the United States government has long held.

I would have expected *Nature* to reserve its regrettable sarcasm for situations in which you and the Bush administration disagreed.

Burton Richter

Stanford University, SLAC MS-80, 2575 Sand Hill Road, Menlo Park, California 94025, USA

Fusion: choose Japan for international balance

Sir—Your Editorial “Time for Japan to shine?” (*Nature* **427**, 763; 2004) offered the world a way out of the present stalemate concerning the selection of a site for the fusion project ITER.

In contrast, the correspondence by P. Vandenplas (*Nature* **428**, 119; 2004) seems counter-productive. By staking Europe’s claim to an “uncontested leading role in fusion”, Vandenplas missed the main point. As your Editorial correctly notes, Japan genuinely yearns for an international project of the size of ITER.

If we look beyond fusion and see the broader context of international science, we can see the need for international balance in large science facilities. Technical experts from each of the ITER parties have already certified both Cadarache and Rokkasho to be acceptable sites for ITER. However, when viewed from the level of international policy-making, it is clear that choosing the Rokkasho site will bring Japan into the international community of science leaders.

I applaud the wisdom of your Editorial and its broad view of international science. This opportune moment in fusion’s history may not last long. Fusion scientists everywhere must agree with your impatient plea, “The time has come to choose a site.” For this to happen, we must certainly look beyond our local concerns and work to find the larger political solution that embraces the scientific aspirations of all nations.

Michael E. Mauel

Department of Applied Physics and Applied Mathematics, Columbia University, 500 W. 120th Street, New York, New York 10027, USA

The physics of crowds

Once social scientists inspired physicists. Now the reverse is true.

Critical Mass: How One Thing Leads to Another

by Philip Ball

William Heinemann: 2004. 644 pp. £25

To be published by Farrar, Straus and Giroux in the US

Steven Strogatz

Psychohistory is a remarkably accurate theory of humans' collective behaviour. It predicts the rise and fall of political movements and empires, and even the sweep of civilization itself. Like the calculations at the root of the insurance industry, it is a fundamentally statistical science, valid only for groups large enough for the vagaries of individuals to average out.

Of course, this theory is the stuff of science fiction, as readers of Isaac Asimov's *Foundation* series will recognize. But in the past decade or two, the dream of psychohistory — although admittedly in a much more modest form — has seduced a motley crew of physicists, complex-systems theorists and social scientists. Borrowing the tools of statistical mechanics and nonlinear dynamics, and armed with unprecedented computer power and huge data sets, they have re-examined some classic puzzles about human group behaviour and its social, economic and political ramifications. These include the causes of boom-bust cycles in the economy; the long-tailed distribution of wealth; the stampedes of panicked crowds; how altruism can emerge in a group of selfish individuals; and the infuriating waves of congestion that can clog traffic on an otherwise open highway. The goal, in short, is to develop a physics of society.

In his fascinating new book *Critical Mass*, Philip Ball tells the story of this research in a comprehensive and often captivating way. But what I liked best about the book is that Ball delves far beyond today's headlines. In particular, I was surprised to learn that this intellectual thread has a long history that is the exact reverse of what is generally supposed. Statistical physics did not spawn social physics; it was the other way around.

In the 1870s, when James Clerk Maxwell was developing the kinetic theory of gases, he knew it would be impossible to track the motion of countless particles ricocheting in a hot gas. But he drew hope from an earlier generation of social statisticians, who had demonstrated that enormous populations of people — who were individually every bit as capricious as gas molecules — could nevertheless display precise regularities. After reading Henry Thomas Buckle's *History of Civilization in England*, which espoused this



Human interest: physicists have tried to explain processes from the action of crowds to the economy.

statistical point of view, Maxwell wrote: "Those uniformities which we observe in our experiments with quantities of matter containing millions and millions of molecules are uniformities of the same kind as those explained by Laplace and wondered at by Buckle arising from the slumping together of multitudes of causes each of which is by no means uniform with the others."

Even the term 'statistics' originated in this movement to quantify social phenomena. In 1749, the German scholar Gottfried Achenwall suggested it as a name for the numerical study of the states of society, which were just beginning to be summarized in terms of death rates, birth rates and other population measures.

Critical Mass is full of historical tidbits like this. Ball particularly enjoys setting the record straight on matters of nomenclature, priority and the like. Thus we learn that Adam Smith argued that a hidden hand, not an invisible one, ruled free markets, and that Maxwell's demon was so named by William Thomson, much to the chagrin of the devoutly religious Maxwell. Maxwell had described his hypothetical creature more lovingly, as "a very observant and neat-figured being".

This book is impressively clear and breathtaking in scope. Ball explains the basics of every subject he treats, from the

political upheavals in seventeenth-century England that gave rise to Thomas Hobbes' vision of utopia, to van der Waals' theory of phase transitions between liquids and gases. Furthermore, unlike so many recent books about complexity theory, this one sets everything in proper historical context, and does not oversell its ideas. Ball even draws on pop literature. For example, to illustrate the abruptness and long-range character of phase transitions, he invokes Kurt Vonnegut's *Grand Ah-Whoom* from *Cat's Cradle*, in which a shard of ice-nine seeds the freezing of all the world's oceans.

On the other hand, I did find Ball's treatment of the discovery process a bit disembodied. Or to say that more positively, this is a book about ideas, not about scientists and their struggles. You won't learn much about how the pioneers in this field made their breakthroughs, or about what drove them, or what they were like. But perhaps that would be asking too much. The book is already massive, and its heft made me wonder if its title were a pun.

Speaking of which, the title *Critical Mass: How One Thing Leads to Another* struck me as uncannily (or perhaps intentionally?) close to the title of Malcolm Gladwell's recent bestseller, *The Tipping Point: How Little Things Can Make a Big Difference*. Both books deal with collective phenomena in

society, or what Gladwell calls social epidemics, but the resemblance ends there. Gladwell's book is fluffy and full of entertaining anecdotes, but is often unconvincing. Ball's is substantial, impeccably researched and generally more persuasive. For anyone who would like to learn about the intellectual ferment at the surprising junction of physics and social science, *Critical Mass* is the place to start. ■

Steven Strogatz is in the Center for Applied Mathematics and the Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853-1503, USA.

Chemistry between man and insect

For Love of Insects

by Thomas Eisner

Harvard University Press: 2003. 464 pp.

\$29.95, £19.95

Jeremy N. McNeil

Chemical ecology, the study of how natural chemicals affect the interactions between organisms, has become a well-recognized discipline in the past 50 years. A number of remarkable scientists have moulded the field, many of whom mix their love of science with an appreciation for the arts and life in general, making them true Renaissance individuals. Thomas Eisner, author of *For Love of Insects*, is one such person: as well as a world-renowned scientist and teacher, he is an excellent photographer and talented musician.

For Love of Insects is not your typical biological text integrating information from the literature on a topic. Instead this book is a series of vignettes. Each of the ten chapters discusses a different chemical ecology project that Eisner and his colleagues have carried out over the past 40 years. The book is well written and beautifully illustrated with colour photographs, the majority taken by the author. I would strongly recommend this book for several reasons.

First, each chapter introduces the reader to different facets of the wonderful world of insects. We learn how many insects defend themselves by spraying aggressors with noxious secretions, how they hit their target, and how they keep themselves from suffering the effects of the compounds they deploy. We discover that insects may produce their own compounds or get them from their food. In some instances, females may receive them from males during mating, and the quantity of a male's defence compounds may determine whether he gets a mate.

Some insects defend themselves with their excrement, or disguise themselves with flower petals. Others adopt a strategy by

which they can feed on aphids while avoiding the ants guarding the aphid colony. We also learn of the ingenious ways in which defences may be overcome, such as the herbivorous insects that can feed on milkweed leaves by severing the veins that carry latex to the leaf, or how some predators eat only part of their prey to avoid toxins. And if one is

on the lookout for research projects, Eisner provides some pointers. For example, what is the role of the second soldier caste in the termite *Nasutitermes exitiosus*? Are blister beetles the only source of cantharidin, a defence chemical that is better known as the pre-Viagra Spanish fly, for cantharidiphilous insects?

Second, the author shows what is needed to produce first-class ecological research. Like Eisner, one must combine keen observation with a desire to understand the phenomena observed. Furthermore, one needs to design elegant and simple experiments to answer 'why?' and 'how?' questions. This underscores the importance of creative thinking — a timely reminder in the post-genomic world that good science is not just a matter of fancy equipment and sophisticated techniques.

Collaboration is another key ingredient. Eisner's research into the bombardier beetle, an insect that defends itself with boiling secretions — rather like a castle's soldiers pouring hot oil on the heads of besiegers — spanned more than 30 years. He credits his graduate students Daniel Aneshansley and



Termites (top) and blister beetles secrete chemicals with very different effects.

Jeff Dean and 17 others who identified and acquired biological material, provided insights into chemical defences and the role of benzoquinones in particular, and worked out how to photograph the spectacular defensive behaviour.

Two collaborators stand out. Eisner has a long-standing association with chemist Jerry Meinwald, who identified many of the compounds involved in the interactions observed. The other is his wife, Maria Eisner, who took the excellent scanning electron micrographs found in this book and in many of their publications. Both share Eisner's love of music, and I have heard Tom and Jerry perform at several scientific meetings.

Throughout the text one is reminded of the pleasure that the author derives from discovery. Anyone reading this book will themselves embark upon a journey of discovery and come to share, if only at arm's length, Tom's love of insects and the wonders of nature. ■

Jeremy N. McNeil is in the Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada.

T. EISNER, CORNELL UNIV.

A sprinkling of stardust

Stellar Alchemy: The Celestial Origin of Atoms

by Michel Cassé (transl. Stephen Lyle)
Cambridge University Press: 2003. 256 pp.
\$30, £19.99

John J. Cowan

Have you ever wondered how humans fit into the cosmos, what our relationship is to the stars, or how the elements were formed? If so, you should enjoy *Stellar Alchemy*. Michel Cassé, a French astrophysicist, takes the reader on an idiosyncratic journey through the intricacies of modern astrophysics to show how the stars, galaxies, the Universe and humans are interrelated. The book, published in French and now translated into English by Stephen Lyle, is designed for general, but scientifically literate, readers.

The book's focus is on nucleosynthesis, the formation of the elements. The Big Bang created hydrogen, helium and some lithium. Two other light elements, beryllium and boron, are made in interstellar space from interactions between cosmic rays and gas atoms. Over the 14-billion-year lifetime of the Universe, all the other elements have been synthesized inside stars and ejected into space to make new stars, planets and even us. In this process a common element, such as iron, could be transformed into gold — the alchemists' dream.

To explain this requires an understanding of the births, evolution and particularly deaths of stars. Cassé draws colourful comparisons between stellar and human life cycles, although he notes that stars are not alive in the same sense as humans. As stars are the primary components of galaxies, which are strewn across the Universe, the formation of galaxies and the origin and fate of the Universe are also described in great and interesting detail, as are the requisite and related physics: the nature of light, the spectra and atomic physics we use to detect elements in stars, and the various aspects of nuclear physics involved in making the elements.

If the focus is on nucleosynthesis, the book's theme is that we are stardust — the atoms in our bodies were once in stars, making stars our distant ancestors. This idea, popularized by the late Carl Sagan, emphasizes the connections between life on Earth and the formation of elements such as carbon and oxygen in stars. In a personal and reflective manner, Cassé places humanity's origins and future squarely in the cosmos. He includes a nice discussion of the death of our Sun, the Earth and the Solar System, and speculates on humans emigrating from



The ancestral home? We are made up of atoms that were created inside distant stars.

Earth before it is too late. Cassé's optimism shines through as he guarantees we will have mastered space flight by that time, assuming we survive as a species.

The book addresses many of the topical questions in astrophysics. There are lengthy discussions and speculations concerning the dark matter that composes more than 90% of the Universe but cannot be seen, and the recently discovered dark energy, an apparent antigravitational effect that is causing an accelerating expansion of the Universe. Although these topics are still mysterious,

Cassé remains the optimist, noting that their explanations may be "unknown but not unknowable".

Cassé's writing style, flowery and poetic and full of historical references, makes even the more technical material accessible to the general reader. The book is also filled with delightful aphorisms, such as "patience is a virtue in the hunt for the invisible" and "the best way of finding a needle in a haystack is to burn the hay". His organization and chapter titles are whimsical, but work — "The Sociology of Stars and Clouds", for

New in paperback

Oxygen: The Molecule that Made the World

by Nick Lane

Oxford University Press, £8.99, \$15.95

"Nick Lane's enjoyable and informative book would have us believe that without the threat of oxygen toxicity, life would never have evolved beyond a green slime. A nicely crafted account of an important element's place in our lives ... deserves to be read widely." Tom Kirkwood, *Nature* **419**, 785 (2002).

Who Owns Academic Work? Battling for Control of Intellectual Property

by Corynne McSherry

Harvard University Press, \$16.95, £10.95, €15.70

What it means to be 98% Chimpanzee: Apes, People, and Their Genes

by Jonathan Marks

University of California Press, \$17.95, £11.95

The Discovery of the Germ

by John Waller

Icon Books, £6.99

Dreaming: An Introduction to the Science of Sleep

by J. Allan Hobson

Oxford University Press, £8.99, \$14.95

A Brain for All Seasons: Human Evolution and Abrupt Climate Change

by William H. Calvin

University of Chicago Press, \$15, £10.50

Divine proportion and the Holy Grail

The idea that the golden section had a special place in Renaissance painting makes for fine fiction but rotten history.

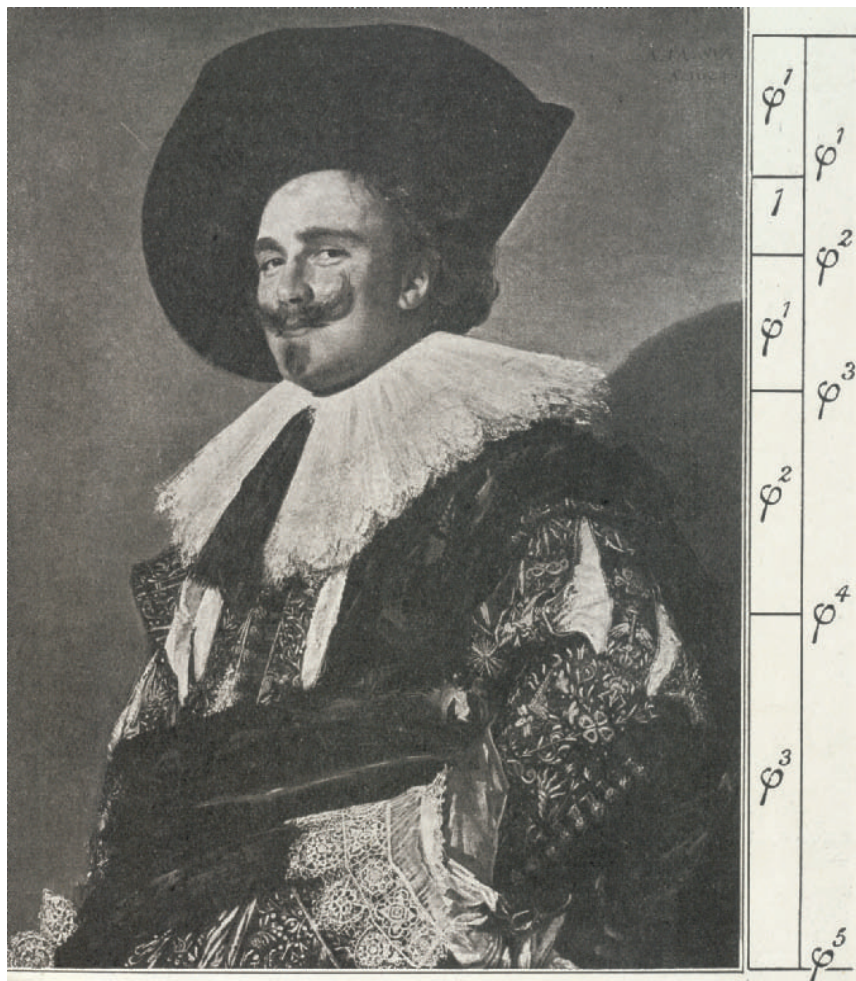
Martin Kemp

Everyone loves a conspiracy. Or so we read in Dan Brown's massive bestseller *The Da Vinci Code*. The conspiracy in this case involves the Priory of Sion, the secret society of which Leonardo was the reputed grand master between 1510 and 1519. The visual secrets include Leonardo's portrayal of St John in the *Last Supper* — young and somnambulant, as was customary — who is re-identified as Mary Magdalene, pregnant with Jesus's child. Mary, as divine mother, is herself the elusive Holy Grail. The linked properties of the 'divine proportion' and the Fibonacci series, beloved of mathematical mystics, provide clues in the decoding of the historical secrets.

Leonardo was certainly aware of the divine proportion (or golden section), illustrating the regular and semi-regular polyhedra in Luca Pacioli's manuscript of the same name in 1498. This section arises when a line is cut so that the ratio of the larger part to the smaller is equivalent to the ratio between the whole and the larger, approximately 1:1.618. The ratio between successive numbers in the Fibonacci series — 1, 1, 2, 3, 5, 8, 13, 21, 34... — approaches this ratio. There is no obvious problem in thinking that Leonardo knew the work of his namesake, the thirteenth-century mathematician Leonardo Fibonacci da Pisa.

The problem lies in granting a privileged significance to the ratio in the Renaissance, or in any period before the later nineteenth century. Incommensurable ratios were regarded as special, but, in Renaissance design, the ratio $1:\sqrt{2}$, based on Pythagoras' theorem, was used far more widely, not least because it was a simple construction based on the diagonal of a square. It features in architectural planning and occasionally to determine the dimensions of paintings.

An ardent advocate of the ubiquity of the golden section in art and nature was the English writer and Olympic swordsman Theodore Andrea Cook. In his *Spirals in Nature and Art* (1903) and *The Curves of Life* (1914), he adduced instances of the 'phi' ratio, which he named after the legendary Greek sculptor Phidias, in such natural configurations as shells, spiral horns and phyllotaxis. He recognized comparable formations in art and artefacts from across the centuries and around the globe. Such was his faith in the ratio that he used it to disclose the aesthetic secrets of some unlikely candidates, including *The Laughing Cavalier* by Frans Hals,



A certain ratio: Theodore Cook thought *The Laughing Cavalier* was based on mathematics.

reproduced here from Cook's 1914 book.

Some artists in the early twentieth century did design their pictures in this way, in particular the Spanish Cubist Juan Gris. But there is no evidence that Renaissance and Baroque artists used such surface geometry in constructing their paintings. Among the thousands of preparatory drawings and the growing number of underdrawings detected in paintings, not one reveals constructional methods based upon the 'secret' division of their surfaces.

There is quite a fashion for the drawing of fat lines on thin reproductions of Renaissance paintings. Draw enough lines on a small image — equilateral triangles, pentagrams and so on — and it is hard not to hit some 'significant' feature. The latest

manifestation of this strategy is Bulent Atalay's surprisingly ahistorical *Math and the Mona Lisa* (Smithsonian, 2004), which is determined to credit Leonardo with Cook-like obsessions.

In the service of fiction, the adducing of mathematical secrets in the quest for the Holy Grail is fine if it works as story-telling. As serious history of image-making in the Renaissance it is nonsense. The problems with Brown's code are not the fantasies and anachronisms, but that his invented truth has been taken seriously by those who cannot recognize fiction.

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

example. Another nice touch is a glossary of technical terms at the beginning of each chapter. Even with those, however, parts of the book may require effort for those without a background in physics and astronomy.

Some of the figures also seem to be more appropriate for, and have come from,

researchers in the field, and some of the tables and appendices are for specialists only. Cassé also stresses the individual and national contributions from French astronomers. While no doubt deserved, their frequent mention seems somewhat out of place. But these are minor quibbles, and Cassé's interesting story

about the relations between the stars and humans, and his joy in the subject, carry the reader along on a pleasant journey. ■

John J. Cowan is in the Department of Physics and Astronomy, University of Oklahoma, 440 West Brooks Road, Norman, Oklahoma 73019, USA.

Cell bodies in a cage

A new 'cell theory' is needed to explain complex intercellular connections.

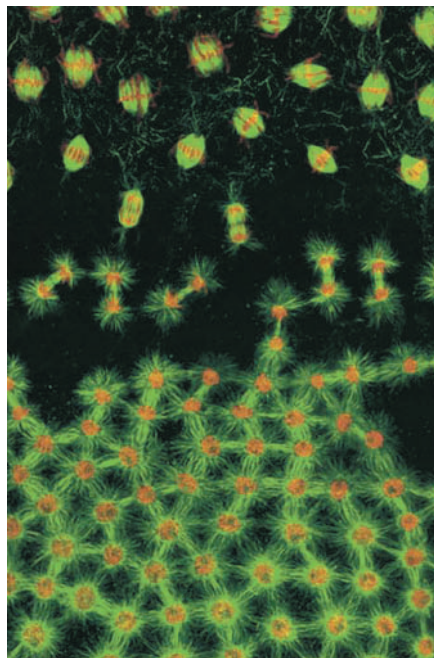
František Baluška, Dieter Volkman and Peter W. Barlow

The concept of the cell as the fundamental structural and functional unit of a multicellular organism stems from the observations of Robert Hooke in 1665 and Nehemiah Grew in 1682, both of whom reported on the 'cells' of plant tissues. It was 150 years before the universal 'German cell theory' — as it was originally known — was proposed by Theodore Schwann and Matthias Schleiden, but even this was not widely accepted for another 100 years. Brain tissues, with their large neurons interconnected by very long thin processes known as axons, had represented a particular hurdle for the early cytologists. But after the cellular basis of neurons and brain tissue had been accepted following the work of Ramón y Cajal and Camillo Golgi, cell theory spread rapidly as a rather dogmatic doctrine.

Plants have been notable and important tools in this process of cellular discovery. Not only did they allow the identification of the actual cells, but they also enabled the discovery of nuclei, plasma membrane, microtubules, mitosis and the cell-division cycle. Paradoxically, recent advances in plant cell biology challenge the dominant position of the current version of cell theory.

Cell theory identifies the cell as the elementary unit from which all living organisms are constructed. Although this holds true for all prokaryotic and unicellular eukaryotic organisms, the supracellular structure of higher plants presents a problem. Cell-to-cell channels, called plasmodesmata, connect each plant cell to its neighbours, facilitating the exchange of large molecules (proteins and RNAs) and allowing the mass flow of smaller molecules. Thus, in contradiction with the cell theory, plant cells are neither physically separated nor structurally independent. The challenge to the cell theory does not stop with plants. There are numerous examples of supracellular assemblages known as coenocytes (which are formed by mitosis with the absence of subsequent cell division) and syncytia (which are formed by cells fusing together), both of which are found throughout the eukaryotic superkingdom. These giant cells contain multiple nuclei.

In 1853, Thomas Henry Huxley noted a similar inconsistency between the structure of some tissues and the German cell theory of his day. Huxley was convinced that cells of multicellular organisms are not anatomically isolated. Accordingly, he could not accept that cells were the elementary units of living



Are nuclei and their surrounding microtubules the fundamental unit of supracellular structures?

organisms. Huxley's conceptual problem with cell theory is still relevant, because supracellularity is nowadays an accepted feature of higher plants. Moreover, this identity crisis of the 'cell' is not simply a problem confined to plants, as nanotubular intercellular bridges are also generated *de novo* between animal cells. These cell-to-cell connections can create complex networks of cytoplasmic continuity that facilitate cell-cell transport, of, for example, endosomal-like vesicles. It seems that algae, fungi, plants and animals have all independently developed both supracellularity and multicellularity.

To harmonize all these diverse cytological observations of eukaryotic forms, a discrete subcellular element is required that will take over from the cell as the fundamental unit, not only of eukaryotic structure, but as a propagule of life itself. One possible candidate is the 'cell body' which was proposed for animal cells by the late Daniel Mazia in 1993. The 'cell body' comprises the nucleus and a set of perinuclear radiating microtubules, and can be regarded as the basic unit of eukaryotic life, being both autonomous and self-reproducing.

The 'cell body' concept is particularly suitable for supracellular plants. Plant nuclei make use of the whole of their nuclear surface to organize radiating microtubules. This feature is prominent in multinuclear coenocytes and syncytia found throughout the eukaryotic superkingdom, where the

radiating microtubules determine the regular spacing of individual nuclei ('cell bodies'). Importantly, 'cell body' microtubules radiate from the organizing sites near or at the nuclear surface into the cytoplasm. This extranuclear scaffold structure provides the structural basis for the well known (but little understood) observation that the volume of the nucleus has a fixed relationship to the size of the cell (the cytonuclear ratio), a principle that underlies the organization of every eukaryotic cell.

The origin of the 'cell body' attracts further speculation. Its dual nature (nucleus and microtubules) may be the result of serial and progressive endosymbiotic events that took place during the early evolution of the eukaryotic cell. In fact, the eukaryotic cell itself is supracellular as it has been developed by the merger of several originally free-living cells. After long discussions, the endosymbiotic origin of mitochondria and plastids is now widely accepted.

Whereas 'cell bodies' cannot arise *de novo*, and are formed only by a structural splitting of pre-existing 'cell bodies' during the highly conserved process of mitosis, the plasma membrane, along with its associated structures and assemblies, can be formed *de novo* through secretory and synthetic activities organized by the 'cell body'. In plants this occurs every time a cell divides at cytokinesis! In addition, the 'cell body' is inherently motile, but is confined within the boundary structure of the plasma membrane with its associated actin cytoskeleton. The active exploration of cytoplasmic space by the 'cell bodies' is a feature that makes them appear akin to active, living organisms trapped within a protective cellular 'cage'; this exploration also underlies the cytonuclear ratio. Altogether, the 'cell body' concept represents a satisfactory and coherent conceptual framework for understanding both supracellularity and cellularity throughout the eukaryotic superkingdom. ■

František Baluška and Dieter Volkman are at the Institute of Cellular and Molecular Botany, University of Bonn, Kirschallee 1, 53115 Bonn, Germany. Peter W. Barlow is at the School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK.

FURTHER READING

- Baluška, F., Volkman, D. & Barlow, P. W. *Trends Plant Sci.* **6**, 104–111 (2001).
 Mazzarello, P. *Nature Cell Biol.* **1**, E13–E15 (1999).
 Oparka, K. J. *Trends Plant Sci.* **9**, 33–41 (2004).
 Richmond, M. L. *Nature Rev. Mol. Cell Biol.* **3**, 61–65 (2001).
 Rustom, A., Saffrich, R., Markovic, I., Walther, P. & Gerdes, H.-H. *Science* **303**, 1007–1010 (2004).

Muscling in on hominid evolution

Pete Currie

A molecular difference in the jaw muscles of human and non-human primates has tantalizing echoes in the fossil record. Was this divergence a central event in the evolution of the skull of modern hominids?

Ever since Bishop Wilberforce famously ridiculed the possibility that man was descended from apes, and T. H. Huxley bravely chose primate ancestry rather than ignorance, the debate over our origins has claimed a special place in evolutionary theory. With the acceptance by most of us that we are indeed a product of natural selection, discussions surrounding the issue have cooled somewhat. But exactly how natural selection acted to produce the modern human form has remained hotly contested.

In the history of this debate, the paper by Stedman *et al.* on page 415 of this issue¹ is of special significance. It describes what may be the first functional genetic difference between humans and apes (Fig. 1). Remarkably, the timing of the appearance of this genetic alteration, or mutation, roughly coincides with the appearance of 'human-like' characteristics in the hominid fossil record, and the authors present convincing arguments as to how the mutation could have been responsible for their acquisition.

The evolution of the genus *Homo* is associated with the appearance of several defining features or traits, which set the species within it apart from our more distant ancestors². Among these morphological innovations is a reduced reliance on powerful masticatory (jaw) muscles as a means of breaking down food. Large jaw muscles are a feature of a number of genera of the family Hominidae. These include living primates and extinct forerunners of *Homo*, such as *Paranthropus* and *Australopithecus*, which are believed to possess many features more allied to extant apes. By contrast, a relatively slight masticatory apparatus is associated with the fossil remains of species within *Homo* and modern humans, and the advent of this morphological transition is closely correlated with a dramatic increase in cranial capacity.

How exactly did this shift in cranial morphology come about? By analysing gene sequences in the human genome, Stedman *et al.*¹ have identified a new member of the class of genes that encode myosin heavy chain (*MYH*) that may be associated with this change. Myosin heavy chains are a critical protein component of the sarcomeres, the 'engine room' of skeletal muscle from which contractile force is derived. There are numerous types of myosin heavy chain specialized for the different muscle-contraction

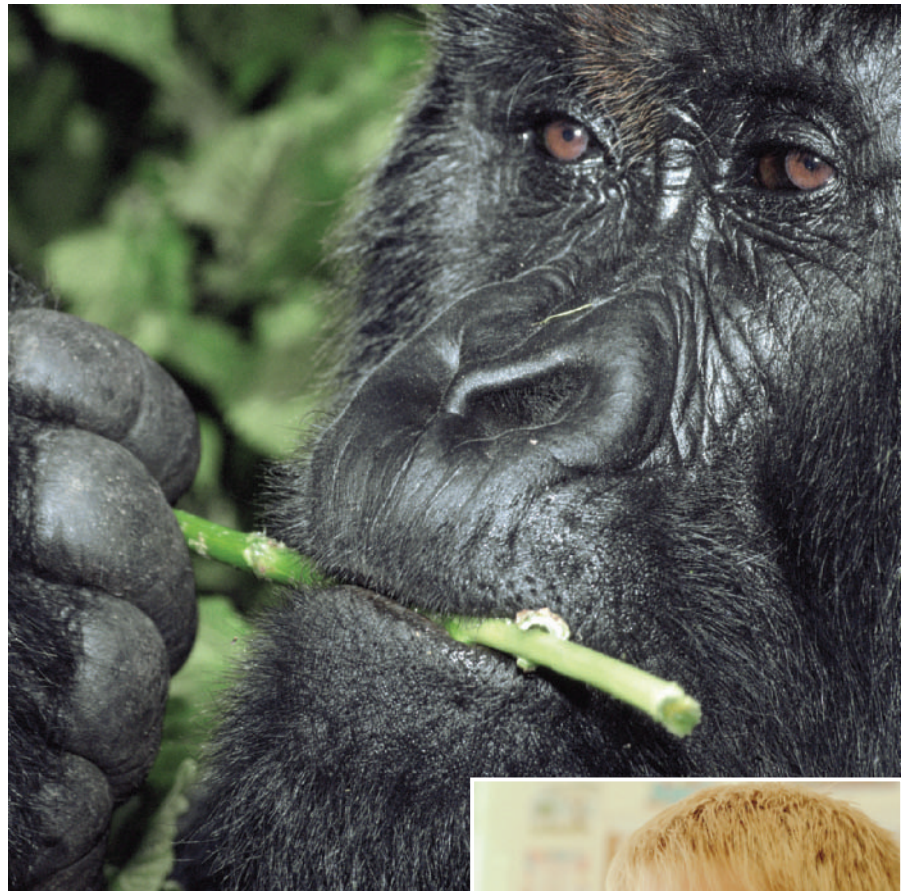


Figure 1 News of chews — the jaw muscles of apes, such as this mountain gorilla, and humans could reflect a profound evolutionary divergence.

rates specific for different muscles. The inactivation of individual *MYH* genes, either in gene 'knockout' mutant mice or in human disease, often results in a dramatic reduction in the size of the muscles in which they are active^{3–5}.

The particular gene in question, *MYH16*, is specifically expressed in the jaw muscles of humans and monkeys. But, surprisingly, a mutation in the human gene prevents the accumulation of MYH16 protein. Stedman *et al.* found that, by contrast, all non-human primates for which genome sequence could be obtained have an intact copy of the gene, and have a high level of MYH16 protein in their jaw muscles. An analysis of the time at which the mutation arose during hominid evolution places it at about 2.4 million years ago, the period just before the evolution of the modern hominid cranial form. These



findings suggest a seductive hypothesis: that a decrease in jaw-muscle size, produced by inactivation of *MYH16*, removed a barrier to the remodelling of the hominid cranium which consequently allowed an increase in the size of the brain.

At first glance, such a proposition looks far-fetched. Surely, such a simple change in muscle anatomy could not alter something as complex and seemingly rigid as the cranial

BRUCE DAVIDSON/NATUREPL.COM

BUTCH MARTIN/ALAMY

skeleton? Surprisingly, however, altering the size of different muscles can produce dramatic alterations in the bones to which they attach^{6,7}. Animal — including primate — models of jaw-muscle transposition or removal show that growth patterns of the craniofacial skeleton can be radically altered by changing muscle anatomy^{8,9}. From these studies it is clear that, over time, the abrupt reduction in masticatory muscle size and contractile force that would have arisen within the *MYH16*-mutant ancestor could have had a considerable impact on cranial morphology. Those effects might well have included a reduction in stress across the bones of the braincase, allowing it to become larger.

Some serious issues, however, are not fully addressed by such a model. The evolutionary acquisition of reduced jaw-muscle size needs explaining in terms of its adaptive significance, independent of any perceived role in craniofacial morphogenesis. Specifically, given the deleterious nature of the *MYH* mutations implicated in human disease, it is unclear how a similar change would have become 'fixed' in the ancestral hominid population. Several explanations could be advanced to counter this ideological roadblock, such as a contemporaneous shift in diet (say, to an increased reliance on meat eating), or a growing dependence on hands rather than the jaw in food preparation. Stedman *et al.*¹ largely ignore the issue, perhaps because it requires a separate, detailed examination. But whatever the immediate consequence of *MYH16* inactivation was, it is now an indicator that a critical change occurred in the hominid masticatory apparatus around 2.5 million years ago.

What is the significance of these findings, and do they shed any light on human origins? Although there is a rough consensus about the individual features that define fossil species within the genus *Homo*, the sequence in which individual traits were acquired during hominid evolution remains controversial. Furthermore, the definition of which character traits were essential for the appearance of the modern human form is equally contentious. The reasons for this are familiar to anyone who tries to explain morphological transitions over large evolutionary distances based primarily on the fossil record. Such explanations hinge on finding so-called 'transitional forms', where a particular fossil is so indelibly etched with the tell-tale signs of what something was, and what it was going to become, that an inescapable evolutionary theory simply tumbles out of the dirt. Not unsurprisingly, such fossils are very rare indeed, and fossils charting the course of hominid evolution are no exception. Stedman and colleagues' identification¹ of the first molecular difference between human and non-human primates, traceable to an anatomical difference

in the fossil record, provides independent evidence and fresh ideas with which to describe the mechanistic basis of hominid evolution.

We can hope that this study¹ represents the vanguard of a new wave of analyses that focus on the genetic basis of human evolution. Any hypothesis about human origins must take into account our understanding of the hominid fossil record. But it is studies such as these that will put meat on the bones of any theory. With the impending completion of the project to sequence the chimpanzee genome, the tantalizing prospect of whole-genome comparisons between humans and our closest living relative is not too far away. It has been suggested that such a comparison could throw up around 40 million nucleotide differences between humans and chimpanzees¹⁰. Identifying which of these differences encode the essential elements of being human is a daunting task. Sophisticated comparative genomic and expression-profile analyses have nonetheless already been completed, revealing

patterns in gene evolution and expression that may guide us to functionally important differences^{11–13}. More than any other report before it, however, the study by Stedman *et al.* suggests that the genetic basis of human evolution can and will be defined. ■

Pete Currie is in the Developmental Biology Program, Victor Chang Cardiac Research Institute, Sydney 2010, Australia.

e-mail: p.currie@victorchang.unsw.edu.au

1. Stedman, H. H. *et al.* *Nature* **428**, 415–418 (2004).
2. Wood, B. *Nature* **355**, 783–790 (1992).
3. Acakpo-Satchivi, L. *et al.* *J. Cell Biol.* **139**, 1219–1229 (1997).
4. Allen, D. L., Harrison, B. C., Sartorius, C., Byrnes, W. C. & Leinwand, L. A. *Am. J. Physiol. Cell Physiol.* **280**, C637–C645 (2001).
5. Martinsson, T. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 14614–14619 (2000).
6. McPherron, A. C., Lawler, A. M. & Lee, S. J. *Nature* **387**, 83–90 (1997).
7. Hamrick, M. W., McPherron, A. C., Lovejoy, C. O. & Hudson, J. *Bone* **27**, 343–349 (2000).
8. Hohl, T. H. *J. Maxillofac. Surg.* **11**, 149–156 (1983).
9. Brennan, M. & Antonyshyn, O. *Plast. Reconstr. Surg.* **97**, 13–24 (1996).
10. Olson, M. V. & Varki, A. *Nature Rev. Genet.* **4**, 20–28 (2003).
11. Caceres, M. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 13030–13035 (2003).
12. Clark, A. G. *et al.* *Science* **302**, 1960–1963 (2003).
13. Wildman, D. E., Uddin, M., Liu, G., Grossman, L. I. & Goodman, M. *Proc. Natl Acad. Sci. USA* **100**, 7181–7188 (2003).

Crystallization

How come you look so good?

Roger J. Davey

Crystallized brown sugar is quite miraculous: it takes just a few 'seeds' in the crystallization process to trigger the formation of many times more crystals. Now we are starting to understand why.

The generation of a solid from a solution through crystallization might sound such a simple, familiar process that one could be forgiven for thinking that it is fully understood. But despite the work of Wilhelm Ostwald¹ on crystal nucleation, and the development of classical nucleation theory², this is not so. Take an everyday example such as brown sugar, millions of tonnes of which are crystallized annually, to be dissolved in tea and coffee. The crystals are all of uniform size — no really big ones, no really small ones — and all are nicely faceted (Fig. 1). This is no accident: the effect is achieved by a process known as 'seeding', in which small crystals of pulverized sugar are introduced in the solution to act as seeds on which crystal growth can start. This seeding, however, does not follow the rules of the garden, with one new plant (or in this case, crystal) per seed. In fact, the amount of seed needed to catalyse the whole crystallization may be less than 1% of the mass of the final product. How do so few seeds give rise to so many crystal nuclei?

We do not really know for sure, and so I am intrigued by the work of Cacciuto and colleagues³, reported on page 404 of this issue. Through computer simulations, Cacciuto *et al.* have investigated just how a

seeding process might work, choosing the case where the seed is a foreign substance rather than a crystal of the solute (as it might be in, for instance, the seeding of clouds to induce rain).

For simplicity and generality, however, Cacciuto *et al.* have focused not on specific molecules but on crystallization in a colloidal suspension of hard, nanometre-size spheres. Using clever simulation methodologies, they have explored how the curvature and size of seeds, both spherical and cylindrical, affect nucleation. When the seeds are hollow spherical sections, varying in radius from 10 to 150 times the diameter of the crystallizing colloidal particles, it seems that nucleation is easier on the convex side of a seed's surface than on its concave side. For cylindrical seeds, it is easier still. This 'experiment' reveals what looks like a 'nucleus factory', centred on a seed particle (see Fig. 4 on page 405), which allows the final number of crystals to be larger than the number of seeds.

Experimentally, the question and application of seeding may go back to Louis Pasteur's discovery in the mid-nineteenth century of optical activity in crystals of sodium ammonium tartrate. The compound's optical activity is due to its chirality, the fact



Figure 1 Hey, brown sugar: Cacciuto *et al.*³ offer new insight into the seeding of crystallization.

that it can exist in two forms, or enantiomers, that are non-superimposable mirror images of each other. In 1866, Desiré Gernez wrote to his former colleague Pasteur, describing the result of an interesting experiment⁴. Following on from Pasteur's work, Gernez had discovered that the addition of seed crystals of pure enantiomer to a racemic solution of the tartrate — one containing equal amounts of the two enantiomers — yielded, not a racemic solid, but crystals of the same chirality as the seed. Separations based on this observation have become known as 'resolutions by entrainment'⁴ and are part of the armoury of the modern-day chemical-process developer.

Not surprisingly, it was chemical engineers, interested in designing continuous crystallization processes, for whom seeding (or secondary nucleation, as they termed it) became a central issue. In 1934, Ting and McCabe⁵ showed that solutions of magne-

sium sulphate could be nucleated more reproducibly at moderate supersaturations in the presence of seeds. Today, commercial crystallization processes operate at suspension densities of perhaps 20%, ensuring that seeding levels are always high.

Some clever experiments⁶ in the 1970s on the seeded nucleation of enantiomers of sodium chlorate revealed that, as long as the supersaturations were not too high, all the crystals were enantiomerically identical to the seed. The experiments also showed that the new crystals originated from the seeds through their contact with the crystallizing vessel. We now know that secondary nucleation, and hence seeding, can often be more effective because of mechanical and liquid-shear damage at the seed surface⁷. Such damage would remove potential nuclei from the seed, allowing them to become free-growing crystals. This seems to be a tantalizing reflection of what Cacciuto *et al.*³ have now shown.

Time, I think, for some new experiments.

Roger J. Davey is at the Molecular Materials Centre, Department of Chemical Engineering, UMIST, Manchester M60 1QD, UK.

e-mail: roger.davey@umist.ac.uk

1. Ostwald, F. W. Z. *Phys. Chem.* 22, 289–302 (1897).
2. Garside, J. & Davey, R. J. *From Molecules to Crystallizers — An Introduction to Crystallization* Ch. 3 (Oxford Univ. Press, 2000).
3. Cacciuto, A., Auer, A. & Frenkel, D. *Nature* 428, 404–406 (2004).
4. Jacques, J., Collet, A. & Wilen, S. H. *Enantiomers, Racemates, and Resolutions* 223 (Wiley, Chichester, 1994).
5. Ting, H. H. & McCabe, W. L. *Ind. Eng. Chem.* 26, 1201–1207 (1934).
6. Denk, E. G. & Botsaris, G. D. *J. Cryst. Growth* 13/14, 493–499 (1972).
7. Mullin, J. W. *Crystallization* 3rd edn, Chs 5, 6 (Butterworth-Heinemann, Burlington, MA, 1992).

RNA interference

Human genes hit the big screen

Andrew Fraser

Genetic screens are powerful tools for identifying the genes involved in specific biological processes. At last, RNA interference makes large-scale screens possible in mammalian cells.

One of the most intuitive ways to learn how a complicated machine works is to take it apart piece by piece — a directed 'learning by breaking'. For biologists, teasing apart the machinery underlying the form and function of an organism can be done, most simply, by removing genes one at a time and looking at the effect. One experimental method for turning genes off is known as RNA interference (RNAi; Box 1, overleaf); this has shot to prominence because it allows almost any gene of known sequence to be shut down with apparently magical ease¹.

In two of the biologist's favourite model organisms, nematode worms and fruitflies, RNAi has been used to turn off almost every

one of their genes^{2,3}. Such genome-wide RNAi surveys of gene function have remained out of reach in mammals — until now, that is, for on pages 427 and 431 of this issue Paddison *et al.*⁴ and Berns *et al.*⁵ report the generation of tools to allow RNAi mass-screening of mammalian genes. This at last makes it possible to carry out genetic screens in mammalian cells in culture.

There is a range of effective strategies for RNAi in mammalian cells (reviewed in ref. 6), and they differ principally in the method for getting the double-stranded RNA (dsRNA) that specifically interferes with the target gene into the cells. In one method, rather than synthesize the dsRNA chemically before introducing it into the cells, the

interfering dsRNA is made directly by the cells themselves. A vector directing the transcription of precise short hairpin RNAs — shRNAs — by RNA polymerase III is introduced into the cells; these transcribed shRNAs are processed by the cell to give the small interfering dsRNAs (siRNAs) that turn off the target gene. shRNA-expressing vectors allow for sustained RNAi in a wide range of cell lines (including embryonic stem cells, the subject of much current research). A complete library of shRNA-expressing vectors designed to target each and every gene in a mammalian genome would thus allow genome-wide RNAi-based genetic screens in cells in culture. Put simply, for any process that we are interested in (cell division, response to DNA damage and so on), with such an shRNA library we could screen every gene in the human genome and ask if it is involved.

Both groups^{4,5} have converged on the same basic shRNA library approach, each generating a retrovirus-based library capable of targeting around a third of human genes; the genes were chosen for their potential roles in disease. Different shRNAs often interfere to differing extents with a target gene, so at least three shRNAs have been cloned for most genes. This multiple coverage not only provides an internal control, but may also allow comparison of both strong and weak 'knock-downs' of a specific gene in an analogous way to a classical genetic approach⁷.

Berns *et al.*⁵ used their library to search for genes that affect the function of *p53*, a tumour-suppressor gene that kills or 'arrests' cells with damaged DNA. They screened around 8,000 human genes to find those required for a *p53*-dependent arrest of cell proliferation and identified six genes, including *p53* itself. Further assays confirm that these genes — which include a histone acetyl transferase and a histone deacetylase, two key regulators of gene expression — do indeed play a role in *p53*-induced cell-cycle arrest and senescence. This ability to survey the gene functions of a full third of the human genome so rapidly is breathtaking, and the success of the subsequent assays underscores the quality of this approach.

The retrovirus-based vectors used by both groups are excellent for many cell-based screens. But they cannot be used for the stable expression of shRNAs in all cell types — this requires moving the shRNA-encoding inserts to different vectors. The shRNA library described by Paddison *et al.*⁴ incorporates an elegant system for shuttling the inserts into any destination vector simply using bacterial mating. Their sequence-verified shRNA library targets almost 10,000 human genes; the shRNAs have also been chosen to allow targeting of the mouse orthologues (equivalents) of those human genes, if possible, and over 5,000 mouse

Box 1 RNA interference: a primer

RNA interference silences a target gene through the specific destruction of that gene's messenger RNA, the intermediary molecule between DNA and protein. Double-stranded RNA (dsRNA) is central to the technique: when dsRNA with identical sequences to a specific mRNA is introduced into cells, the mRNA is recognized and degraded by a multiprotein body called the RNA-induced silencing complex. Destruction of the target mRNA leads to a drop in the levels of its encoded protein, and thus to inhibition

of the target gene.

In worms and flies, dsRNAs of hundreds of nucleotides can be used to target a gene.

However, in mammalian cells long dsRNAs induce a potent anti-viral response, shutting down the synthesis of all proteins. So more sophisticated strategies are required, and small interfering RNAs (siRNAs) are used instead. These siRNAs are about 21 nucleotides long, and are efficiently used by the RNA-induced silencing complex but are too short to activate a full-blown anti-viral dsRNA response.

siRNAs can either be made *in vitro* and subsequently introduced into cells, or they can be made directly in cells through the expression of short hairpin RNAs (shRNAs). shRNAs fold back on themselves, creating a region of dsRNA and a loop. This hairpin is processed enzymatically to remove the loop and generate a mature siRNA. Expression of shRNAs can be used to induce RNAi in transgenic mice as well as in cell lines, so the technique can be applied to investigate gene function in whole animals.

A.F.

genes can be targeted using their clones. This library of easily transferable shRNAs is a beautifully designed resource, and should permit an impressive range of analyses in diverse cell types.

To increase the speed of RNAi screening, both groups^{4,5} borrow a sequence identifier (bar-code) system, developed in studies on yeast, for the quantitative analysis of pools of genes⁸. Each shRNA construct has a unique bar-code — Berns *et al.* use the shRNA sequence itself, whereas Paddison *et al.* have an independent bar-code, which they report as being of far greater effectiveness. The abundance of each shRNA construct in a pool of constructs can be assessed by monitoring the relative levels of each bar-code using a microarray. Thus any screen for genes that confer a growth advantage (or defect) can be carried out by the simultaneous screening of large pools of shRNA-expressing vectors, greatly increasing the throughput. Bar-coding is still in its infancy but has great potential for analysing RNAi selection screens.

There are still some uncertainties surrounding mammalian cell RNAi, especially regarding both specificity and efficiency of targeting. According to one report⁹, a sequence identity of as few as 11–12 nucleotides between an interfering RNA and a messenger RNA may be sufficient for interference to occur. If it is, cross-reactivity is a substantial problem: far from targeting one gene, many expressed shRNAs may target several genes simultaneously. Similar analyses¹⁰ came to the opposite conclusion, however, so it remains to be seen whether this is a general problem. Even if cross-reactivity does occur, there are straightforward controls for specificity: most simply, if two independent shRNAs targeting the same gene give similar effects, it is probably safe to conclude that this is specific to the targeted gene, and not due to some 'off-target' cross-

reaction. This is precisely the approach adopted by Berns *et al.* and the presence in each of the libraries reported here of multiple shRNAs against each gene should make these internal controls relatively easy.

As regards RNAi targeting efficiency, it is clear that — as in worms or flies — different genes in mammalian cells are turned off with differing efficiencies. For example, Paddison *et al.* screened their library to identify components of the proteasome, a cellular machinery that degrades many unwanted proteins and that is implicated in certain diseases. Although genes encoding some subunits (those for the 19S base, for example) were apparently easily identified, others (such as those of the 19S lid or 20S core) were harder to hit. Like any screening tool, RNAi is unlikely ever to be perfect. As the rules for predicting effective shRNAs continue to improve, however, the false-negative rate will drop, and the libraries will improve.

Despite these notes of caution, we will no doubt see an explosion in RNAi screening of mammalian cells over the coming months. As with any genetic screen, the power of each RNAi screen depends on the appropriate choice of functional read-out, and that will require development of a variety of cell-based assays (such as the assay for proteasomal function reported by Paddison *et al.*). As no single laboratory can specialize in every aspect of gene function, the general availability of these shRNA libraries as communal resources is a major step forward, harnessing the screening expertise of the entire mammalian-cell research community. Pulling together the data from these varied RNAi screens in a common, central database will take our understanding of mammalian gene function a further giant stride forward.

Andrew Fraser is at the Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus,



100 YEARS AGO

It is not surprising to find that at last a 'motor' pocket book has appeared; in fact, it is a wonder such a work has not appeared sooner... Our author has a breezy style of expression which adds largely to the pleasure of reading the book. Take, for instance, his treatment of that all-important worry of the motorist, the 'police'.

Mr O'Gorman says, "to pass unchallenged at a speed in excess of the legal limit — a thing which is daily accomplished by carts, hansoms, and even by the London omnibuses on almost every run when the gradients favour them... remember that by sitting upright with a calm face (on a quiet car) you produce no impression of speed except on turning a corner. If you turn a corner without being able to see down the road you are entering at over 20 miles per hour you deserve to be punished. If, however, you stoop forward... jamb your hat over your eyes, screw up your face, stare intently and anxiously, do a great deal of steering with visible swinging of your body, blow your horn in such a manner as to say 'Get out of my way' frequently, instead of pressing it slowly and peaceably, you will invariably be arrested."

From *Nature* 24 March 1904.

50 YEARS AGO

Another statement claimed by Prof. Dingle to be fallacious is connected with an underlying assumption in experimental science; this assumption is that the repetition of an experiment will reproduce the original results. But experimental science is not based on an assumption; "it is an adventure in which you accept whatever you find, and although you may be guided in a particular case by an expectation, the experiment may reveal something totally different". An instance of this is found in the case of Schwabe, who counted sunspots with the object of finding an intra-Mercurial planet, and instead of doing so he found the eleven-year solar period... it would be futile to believe that the achievements of experimental science would necessarily lose all significance if it were discovered that some assumption proved baseless. In the realm of psychology it is accepted that no experiment when repeated produces the original result, and even in physics it has been held for a long time that no experiment is repeatable, the entropy of the universe never being twice the same.

From *Nature* 27 March 1954.

Hinxton, Cambridge CB10 1SA, UK.

e-mail: agf@sanger.ac.uk

1. Fire, A. *et al.* *Nature* **391**, 806–811 (1998).
2. Kamath, R. S. *et al.* *Nature* **421**, 231–237 (2003).
3. Kiger, A. *et al.* *J. Biol.* **2**, 27 (2003).
4. Paddison, P. J. *et al.* *Nature* **428**, 427–431 (2004).
5. Berns, K. *et al.* *Nature* **428**, 431–437 (2004).
6. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. *Nature Rev. Mol. Cell Biol.* **4**, 457–467 (2003).
7. Hemann, M. T. *et al.* *Nature Genet.* **33**, 396–400 (2003).
8. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. *Nature Genet.* **14**, 450–456 (1996).
9. Jackson, A. L. *et al.* *Nature Biotechnol.* **21**, 635–637 (2003).
10. Chi, J. T. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 6343–6346 (2003).

Learning theory

Past performance and future results

Carlo Tomasi

Learning from experience is hard, and predicting how well what we have learned will serve us in the future is even harder. The most useful lessons turn out to be those that are insensitive to small changes in our experience.

A hallmark of intelligent learning is that we can apply what we have learned to new situations. In the mathematical theory of learning, this ability is called generalization. On page 419 of this issue¹, Poggio *et al.* formulate an elegant condition for a learning system to generalize well.

As an illustration, consider practising how to hit a tennis ball. We see the trajectory of the incoming ball, and we react with complex motions of our bodies. Sometimes we hit the ball with the racket's sweet spot and send it where we want; sometimes we do less well. In the theory of supervised learning, an input–output pair exemplified by a trajectory and the corresponding reaction is called a training sample. A learning algorithm observes many training samples and computes a function that maps inputs to outputs. The learned function generalizes well if it does about as well on new inputs as on the old ones: if this is true, our performance during tennis practice is a reliable indication of how well we will play during the game.

Given an appropriate measure for the 'cost' of a poor hit, the algorithm could choose the least expensive function over the set of training samples, an approach to learning called empirical risk minimization. A classical result² in learning theory shows that the functions learned through empirical risk minimization generalize well only if the 'hypothesis space' from which they are chosen is simple enough. That there may be trouble in a poor choice of hypotheses is a familiar concept in most scientific disciplines. For instance, a high-degree polynomial fitted to a set of data points can swing wildly between them, and these swings decrease our confidence in the ability of the polynomial to make correct predictions about function values between available data points. For similar reasons, we have come to trust Kepler's simple description of the elliptical motion of heavenly bodies more than the elaborate system of deferents, epicycles and equants of Ptolemy's *Almagest*, no matter how well the latter fit the observations.

The classical definition of a 'simple enough' hypothesis space is brilliant but technically involved. For instance, the set of linear functions defined on the plane has a complexity (or Vapnik–Chervonenkis dimension³) of three because this is the greatest number of points that can be arranged on the plane so that suitable linear functions assume any desired combination of signs (positive or negative) when evaluated at the points. This definition is a mouthful already for this simple case. Although this approach has generated powerful learning algorithms², the complexity of hypothesis spaces for many realistic scenarios quickly becomes too hard to measure with this yardstick. In addition, not all learning problems can be formulated through empirical risk minimization, so classical results might not apply.

Poggio *et al.*¹ propose an elegant solution to these difficulties that builds on earlier intuitions^{3–5} and shifts attention away from the hypothesis space. Instead, they require the learning algorithm to be stable if it is to

produce functions that generalize well. In a nutshell, an algorithm is stable if the removal of any one training sample from any large set of samples results almost always in a small change in the learned function. *Post facto*, this makes intuitive sense: if removing one sample has little consequence (stability), then adding a new one should cause little surprise (generalization). For example, we expect that adding or removing an observation in Kepler's catalogue will usually not perturb his laws of planetary motion substantially.

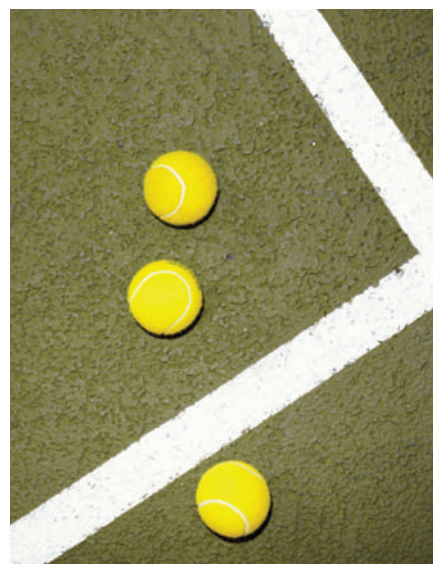
The simplicity and generality of the stability criterion promises practical utility. For example, neuronal synapses in the brain may have to adapt (learn) with little or no memory of past training samples. In these cases, empirical risk minimization does not help, because computing the empirical risk requires access to all past inputs and outputs. In contrast, stability is a natural criterion to use in this context, because it implies predictable behaviour. In addition, stability could conceivably lead to a so-called online algorithm — that is, one that improves its output as new data become available.

Of course, stability is not the whole story, just as being able to predict our tennis performance does not mean that we will play well. If after practice we play as well as the best game contemplated in our hypothesis space, then our learning algorithm is said to be consistent. Poggio *et al.*¹ show that stability is equivalent to consistency for empirical risk minimization, whereas for other learning approaches stability only ensures good generalization. Even so, stability can become a practically important learning tool, as long as some key challenges are met. Specifically, Poggio *et al.*¹ define stability in asymptotic form, by requiring certain limits to vanish as the size of the training set becomes large. In addition, they require this to be the case for all possible probabilistic distributions of the training samples. True applicability to real situations will depend on how well these results can be rephrased for finite set sizes. In other words, can useful measures of stability and generalization be estimated from finite training samples? And is it feasible to develop statistical confidence tests for them? A new, exciting research direction has been opened.

Carlo Tomasi is in the Department of Computer Science, Duke University, Durham, North Carolina 27708, USA.

e-mail: tomasi@cs.duke.edu

1. Poggio, T., Rifkin, R., Mukherjee, S. & Niyogi, P. *Nature* **428**, 419–422 (2004).
2. Vapnik, V. N. *Statistical Learning Theory* (Wiley, New York, 1998).
3. Devroye, L. & Wagner, T. *IEEE Trans. Information Theory* **25**, 601–604 (1979).
4. Bousquet, O. & Elisseeff, A. *J. Machine Learning Res.* **2**, 499–526 (2002).
5. Kutin, S. & Niyogi, P. in *Proc. 18th Conf. Uncertainty in Artificial Intelligence*, Edmonton, Canada, 275–282 (Morgan Kaufmann, San Francisco, 2002).



In or out: success rests on learning algorithms that are stable against slight changes in input conditions¹.

Earth science

Putting the squeeze on oxidation

Michael J. Walter

The results of experiments conducted under the extreme temperature and pressure conditions of Earth's lower mantle suggest that, at these depths, low oxygen content is no barrier to iron oxidation.

The oxidation of iron is a process with which we are all familiar. The surface of Mars, the rust on our cars, the blood in our veins — all bear the conspicuous red colour that marks the oxidation of iron from its metallic (Fe), or divalent (Fe^{2+}), form to a trivalent cationic species (Fe^{3+}). At atmospheric pressure, the process of iron oxidation in the presence of oxygen depends on the partial pressure of gaseous oxygen (often referred to in geological literature as oxygen fugacity). Earth scientists have used the same thermodynamic formulations, with great success, to calibrate the oxidation state of iron in the rocks and minerals of Earth's upper mantle, at much higher pressures and temperatures¹. Now Frost *et al.*² (page 409 of this issue) show that, when it comes to the more extreme pressure conditions in the lower mantle, we must abandon our conventional thinking about oxidation state.

Geology students are taught that certain minerals — such as $(\text{Mg,Fe})\text{Al}_2\text{O}_4$, or 'spinel' — serve as sensors of oxidation state because their $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratios are known, experimentally and thermochemically, to depend on oxygen partial pressure. From such 'geobarometry', we have learnt that Earth's upper mantle is much too oxidized to be in chemical equilibrium with metallic iron¹ — the abundance of Fe^{3+} is too great for metallic iron to be stable. In contrast, we are quite confident that the mantle as a whole was in chemical equilibrium with metallic iron as the planet's core formed (when Earth was less than 50 million years old)³. How the uppermost mantle, the part we can sample, became so oxidized is unknown.

The portion of the mantle at depths greater than 660 km, referred to as the lower mantle, is thought to be dominated (70–80% by weight) by a dense, magnesium–iron silicate, $(\text{Mg,Fe,Al})(\text{Si,Al})\text{O}_3$. This mineral has a stable 'perovskite' crystal structure (a structure well known in ceramics to have an array of rich and unusual properties). The oxygen content of the lower mantle is unknown because we have no direct rock samples, but it is usually assumed to be similar to that of the upper mantle. In 1997, McCammon⁴ reported the surprising discovery that magnesium–silicate perovskites synthesized in the laboratory can have very high $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratios — much higher, for example, than are preserved by spinel in the relatively oxidized upper mantle. However, the relationship between the oxygen

partial pressure and the $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratio in perovskite remained unclear because of the difficulty of knowing (let alone controlling) the oxygen content in experiments at extreme conditions.

Frost *et al.*² have performed experiments in which they brought magnesium perovskite into chemical equilibrium with metallic iron under pressure–temperature conditions typical of the uppermost lower mantle. They made the remarkable discovery that, with very low oxygen content in their samples, the perovskite synthesized can have surprisingly high $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratios — similar, in fact, to the ratios measured in perovskites synthesized at very high oxygen contents. This decoupling of oxygen content from the $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratio shatters our traditional view of oxidation state based on oxygen partial pressure.

This result leads Frost *et al.*² to predict that a free iron-metal phase should exist in Earth's lower mantle — only 1% by weight, but nevertheless an important component. Ironically, this metal phase forms because of the high $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratio in perovskite, rather than being excluded by it. For a mineral to achieve a high $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratio, oxygen must be available to convert FeO into Fe_2O_3 (Fe^{2+} to Fe^{3+}). This could happen in the mantle if some chemical reaction serves as an oxygen pump: for example, the reduction of CO_2 to form C and O_2 . An alternative mechanism

proposed by Frost *et al.* is that magnesium perovskite has such an affinity for Fe^{3+} that it is energetically favourable for FeO in perovskite to 'self-reduce' by the reaction $3\text{Fe}^{2+}\text{O} \rightarrow \text{Fe}(\text{metal}) + 2\text{Fe}^{3+}\text{O}_{1.5}$. In this case, no external oxygen pump is needed and a free iron-metal phase forms out of the necessity for mass balance.

The lower mantle constitutes a large volume of the Earth, about 58%, so even a small amount of metallic iron represents a substantial reservoir — several per cent of all metallic iron in the planet. The lower mantle may first have formed because of the increase in pressure as the solid Earth grew, or it may have crystallized from a deep magma ocean. The core formed very early in Earth's history by the segregation of metal from silicate, a large-scale differentiation event that substantially decreased the Fe/O ratio of the silicate mantle. This ratio, dictated by metal–silicate equilibrium, would have been considerably higher than the current Fe/O ratio in the upper mantle. Frost *et al.*² calculate that if about 10% of the perovskite-manufactured free-metal phase were mechanically removed from the lower mantle through interaction with descending plumes of iron during core segregation, the mantle could have acquired its present Fe/O ratio. Thus, Frost and colleagues' discovery may explain the vexing question of why the upper mantle seems to have had such a high oxidation state for the past 4 billion years⁵.

Michael J. Walter is at the Institute for Study of the Earth's Interior, Okayama University, Misasa, Tottori 682-0193, Japan.

e-mail: walter@misasa.okayama-u.ac.jp

1. O'Neill, H. S. C. & Wall, J. V. J. *Petrol.* **28**, 1169–1191 (1987).
2. Frost, D. J. *et al.* *Nature* **428**, 409–412 (2004).
3. Walter, M. J. *et al.* in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 265–290 (Univ. Arizona Press, Tucson, 2000).
4. McCammon, C. *Nature* **387**, 694–696 (1997).
5. Canil, D. *Nature* **389**, 842–845 (1997).

Medicine

Profile of a tumour

Olli Kallioniemi

Molecular profiling — the comprehensive analysis of genes, RNAs and proteins — is having a radical effect on our understanding of cancer. The next step is to convert these findings into better diagnosis and treatment.

The scale of cancer research projects has increased enormously over the past decade or so, presenting scientists with the opportunity to investigate the role of all human genes in cancer. This change has been wrought largely by advances in high-throughput technologies for analysing the human genome, transcriptome and proteome — genes, messenger RNAs and proteins — and in bioinformatic and statistical methods for exploring the gigabytes of data that are generated. As more is learnt

about the genes and signalling pathways that promote the proliferation of cancer cells, prevent their death or allow their spread to other organs, it should become feasible to develop new methods of diagnosis, as well as treatments that are tailor-made to fit the specific molecular patterns of individual patients. But realizing these ideas is proving a challenge, as a recent meeting* made clear.

In one approach for converting the

*CNIO Symposium: The Molecular Taxonomy of Cancer. Madrid, Spain, 3–6 February 2004.

knowledge obtained from 'omics' techniques into clinical application, researchers are attempting to use microarrays — a means of analysing patterns of gene expression by looking at the mRNAs present — to determine the appropriate treatment for different patients. For instance, microarray profiling of breast tumours, using a set of 70 informative genes, has revealed gene-expression signatures that are associated with a good or a poor prognosis^{1,2}. In these studies, roughly 40% of early-stage breast cancers turned out to have a 'good' signature, associated with only a 15% risk of metastasis (the spread of tumour cells) and a 5% risk of death by 10 years after diagnosis. According to current clinical practice, most patients with early-stage breast cancer receive adjuvant therapy (chemotherapy or endocrine therapy after surgery). But less than 1% of the patients with a 'good prognosis' signature would be likely to benefit, and the treatments often have harmful side effects (discussed by L. Van't Veer, Netherlands Cancer Institute, Amsterdam)². So researchers now suggest using the 'poor' signature to guide the administration of adjuvant therapy to those for whom it would be most useful.

Indeed, gene-expression profiling is already being introduced clinically as a diagnostic procedure in academic centres in the Netherlands (L. Van't Veer), with similar programmes being launched in the United States. The profiles have been validated retrospectively in several cohorts of patients. Prospective validation, however, can only be achieved after 5–10 years. Also — as several

speakers discussed — some technical issues have yet to be resolved, including how to procure and process samples and to ensure reproducibility and quality control. Nevertheless, various analysis platforms and short lists of diagnostic genes are being developed to predict the prognosis associated with other types of tumours, such as lymphomas (L. Staudt, NCI, Bethesda, and M. Piris, CNIO, Madrid), or to predict the likelihood of particular endpoints, such as metastasis (T. Golub, Broad Institute, Cambridge, Massachusetts). Predicting a patient's response to specific types of treatment would, of course, be the most important application. But it may take years for the approach to meet with regulatory approval and be extended from academic centres to mainstream practice.

Another strand of research involves the proteomic profiling of blood serum. This could soon complement — if not replace — the use of established tumour 'markers' in cancer screening and early diagnosis (E. Petricoin, FDA, Bethesda). Whereas specific tumour markers are often used singly, proteomic serum profiling involves the use of mass spectrometry to identify up to 15,000 peaks, representing proteins and protein fragments that are defined by their mass-to-charge ratios³. The sensitivity and specificity of such profiles has exceeded 90% for the diagnosis of lung cancer, and approached 100% for ovarian cancer (E. Petricoin). Most of the peaks have still to be identified, and several investigators argued that panels of specific immunoassays, which use antibodies that recognize particular proteins, should

be developed for diagnosis instead. But it is the fragments, not intact proteins, that are often the most informative diagnostically, and there may be dozens — if not hundreds — of such fragments, possibly existing in complexes with one another and with other serum proteins. That makes the development of specific assays demanding.

A further application of proteomics concerns the analysis of signalling pathways, for example by studying the levels of specific phosphorylated proteins — a key feature of many pathways — in tumour samples (E. Petricoin). Analysis of these signalling profiles is being incorporated into many clinical trials to identify biological markers that indicate which tumours are likely to respond to treatment. Proteomic tumour profiling is often carried out from specific cell types, acquired by microdissection of the tumours. Such studies could also help us to understand the tumour microenvironment, where many different cell types and interactions may affect tumour behaviour. For example, in breast cancer, characteristics of the stromal fibroblast cells adjacent to the tumour might change, and contribute to, cancer progression (R. Weinberg, MIT) — much as does the growth of new blood vessels.

Moving on to studies of gene sequence (genomics), mutations in roughly 1% of human genes (291 genes in total) were reported to contribute to cancer (M. Stratton, Sanger Centre, Hinxton)⁴. Twenty-seven of these genes, more than would be expected by chance, encode protein kinases — enzymes that phosphorylate other pro-

Planetary science

Stardust's comet memories

In January this year, NASA's Stardust spacecraft flew within 237 km of the comet Wild2. Stardust took 72 close-up shots during the flyby, described as an "unqualified success" at the Lunar and Planetary Science Conference in Houston, Texas, last week. After a few months' analysis, the Stardust team has now released stunning pictures of the comet's 5-km nucleus.

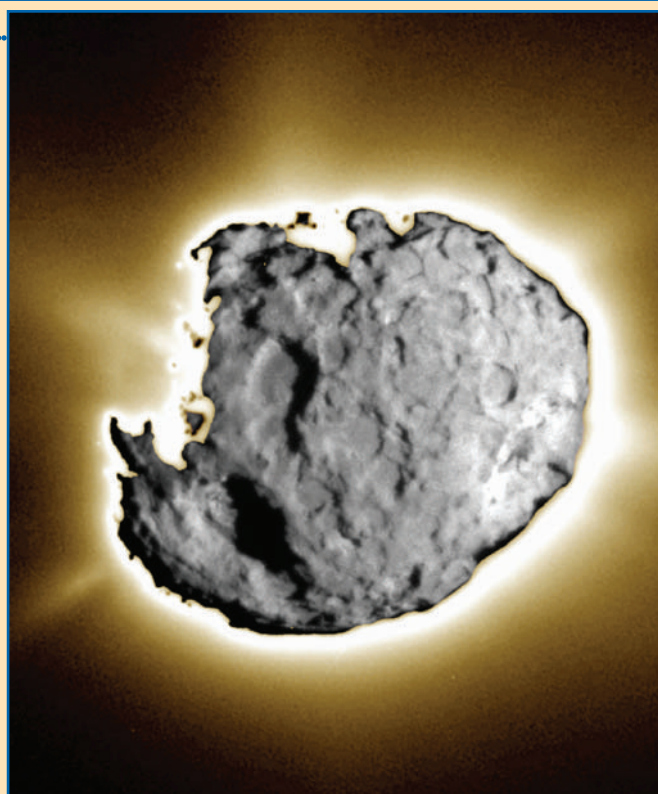
The Wild2 comet is 'fresh'. It is believed to have spent billions of years in the Kuiper belt, out beyond Neptune, until gravitational encounters kicked it into orbit closer to the Sun. In 1974, a close encounter with Jupiter knocked Wild2 into its present orbit, inside that planet's. Since then, Wild2 has passed close to the Sun only five times and so has suffered less exposure to solar radiation than other comets. Its surface bears a

well-preserved record of its early history in the Kuiper belt.

The short-exposure image at the centre of this composite shows the cometary surface. The surface is markedly different from those of the comets Halley and Borrelly, and of Jupiter's ice-rich moons Callisto and Ganymede: instead of a continuum of crater size, Wild2's craters are mostly large; some are likely to be impact craters, others not. A long-exposure image taken 10 seconds later reveals a glowing halo of gas, thrown out from the nucleus in as many as 20 highly collimated jets.

Stardust's mission doesn't end there. The spacecraft is now heading for Earth once more, carrying particles from the comet's coma that were trapped during the flyby in centimetre-size aerogel cells. The samples will arrive on 15 January 2006.

Alison Wright



NASA/JPL

teins. As several approved cancer drugs are inhibitors of tumour-associated kinases, such as BCR/ABL, KIT, HER-2 and the epidermal growth factor receptor, the search for mutated kinases is important. Another mutated kinase is BRAF, identified in a large-scale DNA-sequencing project to be altered in melanomas (reviewed in ref. 4). But an ongoing project to sequence the genes encoding 478 kinases in a panel of breast cancers has yet to replicate the success of the BRAF example (M. Stratton).

A new way to discover the roles of particular genes in cancer is to use cultured cancer cells and the technology of RNA interference to decrease gene expression. Several academic centres and corporations are developing libraries of RNA interference reagents that reduce the expression of any of the 30,000 human genes (G. Hannon, Cold Spring Harbor Laboratory; see also page 375). These libraries will make it possible to identify the functions of genes that control crucial properties of cancer, and could therefore be attractive drug targets. Moreover, several other approaches to the development of cancer treatments are being launched. When specific

target proteins are not known, gene-expression profiles may be used to screen for drugs that modulate cancer-cell behaviour (T. Golub)⁵. For example, the expression pattern of five genes distinguishes leukaemic cells from normal blood cells and can be used as a readout to search for drugs that induce leukaemic cells to differentiate⁵.

Roughly 500 new anticancer drugs are in clinical trials, and many more should soon be under development as a result of new technologies. At the same time, advances in the molecular profiling of tumours could help to define diagnostic patterns that identify those patients most likely to benefit from the new treatments. Although numerous challenges still lie ahead, the future of personalized cancer treatment looks promising. ■

Olli Kallioniemi is at the Medical Biotechnology Unit, VTT Technical Research Centre of Finland, and the University of Turku, FIN-20521 Turku, Finland.

e-mail: oli.kallioniemi@vtt.fi

1. van't Veer, L. J. *et al.* *Nature* **415**, 530–536 (2002).
2. Van de Vijver, M. J. *et al.* *N. Engl. J. Med.* **347**, 1999–2009 (2002).
3. Alexe, G. *et al.* *Proteomics* **4**, 766–783 (2004).
4. Futreal, P. A. *et al.* *Nature Rev. Cancer* **4**, 177–183 (2004).
5. Stegmaier, K. *et al.* *Nature Genet.* **36**, 257–263 (2004).

Signal processing

Neural coding by correlation?

Arun V. Holden

Noise limits the efficiency of information transfer. But correlations of the intervals between signal pulses can reduce low-frequency noise and thereby increase the transfer of information.

A nerve cell is an example of a system whose excitation codes and transmits dynamic information. When a stimulus is above a given threshold, the neuron responds by generating an action potential, so that over time it fires irregular sequences of all-or-nothing spike responses. The ionic cell mechanisms that generate the spike are understood, but there are many mechanisms for coding the spike train, and our knowledge of these mechanisms ranges from well established to speculative. In *Physical Review Letters*, Chacron *et al.*¹ show, using a simple action-potential model, that correlations between sequential interspike intervals can shape the noise spectrum — and that this shaping can increase the transmission of information, because it reduces the noise spectrum at low frequencies.

Information theory is essentially linear²: a reaction is directly proportional to an action. So it might be expected that the constraints imposed by interval correlations would reduce the transmission of information. However, spike generation is a strongly nonlinear, excitable process with a threshold, and such systems can behave counter-intuitively. A good example is stochastic

resonance³, in which additive noise can increase, rather than decrease, the efficiency of information transmission.

A spike train can be represented by the sequence of intervals between spikes; this is characterized by the interval statistics (in the time domain by probability distributions and correlations, and in the frequency domain by spectral densities). Chacron *et al.*¹ consider two schemes of spike generation. The first produces a 'renewal process' that has no memory of the excitation because the system resets itself each time a spike is generated. Here, there is no correlation between successive spike intervals. The probability distributions for higher-order intervals (say, between one spike and the third spike following) and the 'autocorrelation' (the probability of a spike occurring after some other spike, irrespective of how many intervening spikes there were) can be calculated directly from the interspike-interval probability density.

The second scheme generates a non-renewal spike train, with correlations between adjacent intervals. Spike trains of sensory neurons⁴ with a constant stimulus often show dependencies between neighbouring intervals; such dependencies are

produced either by fluctuations in an oscillatory spike-generator or by activity-dependent changes in threshold. Long intervals followed by short ones (and vice versa) result in negative interval correlations.

The essence of spike generation is that inputs too weak to trigger a response are summed, giving a potential; when the potential reaches a threshold, a spike is generated. Chacron *et al.*¹ use a 'perfect integrator model'⁵ for both spike-generation schemes, with a random threshold drawn from a uniform distribution. When the threshold is reached the potential is reset, either by an amount dependent on the magnitude of the threshold just reached (to produce serial correlation), or by a random amount (to generate a spike train with the same interspike-interval density but no correlation).

The advantage of using this simple model is that the spectra, coherence and information transmission rates (mutual information rates) can all be calculated, as well as estimated through computer simulation. Chacron *et al.* demonstrate that a negative serial correlation between spikes reduces the spike-train spectrum and coherence at very low frequencies, although at middle-range frequencies the coherence is increased. Serial correlation also enhances the ability of the model to transmit information by reducing low-frequency noise compared with the renewal process. Such effects have been seen in more realistic neural models⁶, and can be quantified in real neural spike trains; but the complexity of both of these situations had meant that the mechanism for improved information transfer was unclear.

Whether or not this increase in the information transmission rate is exploited in neural systems is an open question, as biological evolution produces systems that work well enough and are robust, without necessarily being optimally efficient. The nervous system responds to a spike train in real time and does not process it as an indefinite sequence. But the effect of correlation on the transmission rate in a single spike train might transfer to correlations between multiple spike trains: variability between different neurons could be correlated, through common inputs and feedback⁷, and coupling within a population could lead to a similar reduction in variability by noise shaping⁸. ■

Arun V. Holden is in the School of Biomedical Sciences, University of Leeds, Leeds LS2 9JT, UK.
e-mail: arun@cbiol.leeds.ac.uk

1. Chacron, M. J., Lindner, B. & Longtin, A. *Phys. Rev. Lett.* **92**, 080601 (2004).
2. Borst, A. & Theunissen, F. E. *Nature Neurosci.* **2**, 947–957 (1999).
3. Jaramillo, F. & Wisenfeld, K. *Nature Neurosci.* **1**, 384–388 (1998).
4. Ratnam, R. & Nelson, M. E. *J. Neurosci.* **20**, 6672–6683 (2000).
5. Stein, R. B., French, A. S. & Holden, A. V. *Biophys. J.* **12**, 295–322 (1972).
6. Chacron, M. J., Longtin, A. & Maler, L. *J. Neurosci.* **21**, 5328–5343 (2001).
7. Azouz, R. & Gray, C. M. *J. Neurosci.* **19**, 2209–2223 (1999).
8. Marr, D. J., Chow, C. C., Gerstner, W., Adams, R. W. & Collins, J. J. *Proc. Natl Acad. Sci. USA* **96**, 10450–10455 (1999).

Archaeology

Buried 'butter' brought to light

Analyst 129, 270–275 (2004)

People digging up slabs of peat for fuel in Ireland and Scotland sometimes find lumps of white, waxy material embedded in the bog. This 'bog butter' seems to be the remains of fatty foods that were buried between about 400 BC and the Middle Ages, to preserve them in the cool anaerobic conditions.

Robert Berstan *et al.* have now analysed the fat molecules in bog butter to figure out its origin. They say that some of it originates from dairy products such as real butter, whereas some derives from animal fatty tissue. In both cases, the material is transformed over time to a waxy substance similar to adipocere, which is found in human corpses buried or recovered from the sea.

The precise composition of the fatty compounds — the length of their carbon chains, and their ratio of carbon-12 to carbon-13 — preserves a signature of their origin, which the researchers discerned using chromatography and mass spectrometry. Out of nine Scottish bog butters, they found that six come from dairy products and three from animal fat, showing that bog burial was used to preserve (or perhaps modify) diverse foodstuffs in ancient times.

Philip Ball

Cancer

Communication breakdown

J. Cell Sci. doi:10.1242/jcs.01000 (2004)

It is generally assumed that cancers start when carcinogens, such as harsh chemicals, create mutations in a cell's DNA that trigger uncontrollable proliferation. For one case at least, Maricel V. Maffini *et al.* challenge that view. They show that cells may become cancerous only after neighbouring tissues, rather than the cells themselves, have been exposed to carcinogens.

The researchers studied cultured tissues of the rat mammary gland, a model of breast cancer in which malignancies are thought to arise in the epithelium rather than the adjacent stroma. They transplanted epithelial cells that had been exposed to either saline or a chemical mutagen into stromal fat tissue that had also been exposed to either saline or a chemical mutagen.

Maffini *et al.* found that healthy epithelial cells transplanted into the chemically treated fat tissue became cancerous, regardless of whether the cells had been mutated themselves. Conversely, chemically treated epithelial cells grew into apparently normal mammary gland ducts when injected into healthy stromal tissue.

The authors suggest that the carcinogen wrecks normal communication between stromal and epithelial cells, and that this encourages epithelial cells to start growing into tumours. The finding implies that other cancers may also form because of a breakdown in signalling between adjacent tissues.

Helen Pearson

Behavioural genetics

Good relations

Curr. Biol. 14, 510–513 (2004)

Western lowland gorillas (*Gorilla gorilla gorilla*) are unusual in that males from neighbouring social groups often mingle without aggression. One explanation for this peaceful coexistence, stemming from DNA analyses, is that the males are related to one another.

Brenda J. Bradley and her colleagues took DNA samples from hair and dung at gorilla nest sites over a 50-km² area of central Africa. They identified 14 silverbacks (mature males), only two of which were unrelated to any of the others analysed. Furthermore, males that lived closer together tended to be more closely related — a pattern not seen in females.

Western gorilla groups usually contain just one silverback, with other males emigrating a short distance as they mature. This, the authors argue, is the basis of their neighbourly behaviour, and is similar to the short-distance male emigrations that characterize marriage in many human societies. Groups of mountain gorillas (*Gorilla gorilla beringei*), on the other hand, often feature several silverbacks and little emigration, which may explain the more frequent fights between males of rival groups.

Michael Hopkin

Plant biology

Autumn's genes

Genome Biol. 5, R24 (2004)

The size and slow growth of trees make them a challenging subject for molecular biologists. Nevertheless, Anders Andersson *et al.* have applied a genomic approach to the study of aspen (*Populus tremula*), and in so doing have provided a molecular picture of the changes that a tree undergoes as winter approaches.

The authors constructed microarrays with which to quantify the genes that were expressed in leaves from a single wild aspen tree during five weeks in August and September. Aspens are deciduous, so they lose their leaves in autumn. During this process, many valuable chemicals — such as the green pigment chlorophyll — are recycled for use in following years, leaving behind brown and gold carotenoid and flavonoid pigments.

Andersson and colleagues' data show the



B. THARP/CORBIS

wide-ranging shift in gene expression that underlies these events. For instance, more than one week before any outward signs appear, protein production increases as leaves start making proteases and other degradative enzymes. Moreover, the production of enzymes involved in photosynthesis ceases, and enzymes that generate energy by burning fats and sugars are made instead. These data provide a map of autumn's molecular fluxes — and also identify 35 genes expressed nowhere else but in autumn leaves.

Christopher Surridge

Organic chemistry

Art or science?

J. Chem. Inf. Comput. Sci. doi:10.1021/ci030415e (2004)

Chemists often claim that the synthesis of organic molecules is more of an art than a science, because of the creativity needed to design a sequence of reactions to produce a precise, complex molecular architecture. Their designs are guided by retrosynthetic analysis — thought experiments that break complicated molecules into simpler units, which serve as starting materials for a real synthesis. The rules of retrosynthesis, derived through experience, always strive for the simplest starting materials within the constraints of what is practically possible.

Christoph Rücker *et al.* have now found that these rules are underpinned by mathematical bedrock. They have used different numerical indicators of molecular complexity — such as how the atoms in a molecule are connected — to plan a range of chemical syntheses. These simple mathematical rules give very similar answers to the heuristic rules of retrosynthesis, and can also rank different options according to how much they simplify the synthesis. The authors conclude that this seemingly artistic part of chemistry is actually deeply scientific.

Rücker *et al.* admit that there are many aspects of chemistry that are still beyond the reach of simple indices. But they recall the words of Immanuel Kant, written in 1786: "In any particular natural philosophy there is only as much real science as there is mathematics to be found in it."

Mark Peplow

Risk of horses falling in the Grand National

Analysis of past tumbles in this gruelling steeplechase points to ways of making it safer.

As in other competitive sports, the famous Grand National steeplechase, which is held at Aintree in the United Kingdom and is watched by 600 million people worldwide¹, sometimes results in injury. By analysing data from the past 15 Grand National races (consisting of 560 starts by horses), we are able to identify several factors that are significantly associated with failure to complete the race: no previous experience of the course and its unique obstacles, unfavourable ground conditions (too soft or too hard), a large number of runners, and the length of the odds ('starting price'). We also find that there is an increased risk of falling at the first fence and at the jump known as Becher's Brook, which has a ditch on the landing side. Our findings indicate ways in which the Grand National could be made safer for horses and illustrate how epidemiological analysis might contribute to preventing injury in competitive sport.

The race comprises 30 obstacles over a distance of 4.5 miles (7.2 km) (Fig. 1). We used multivariable models to identify the variables associated with non-completion and with falling (see supplementary information).

The probability of remaining in the race (Fig. 2a) shows a marked decline at fence 1, followed by a steady decline throughout the rest of the race. Non-completion is further subdivided into different types (Fig. 2b), and demonstrates an increased rate of falling and of horses being 'brought down' early in the race by interference from another falling horse, followed by a higher rate of refusals and horses 'pulling up' (jockey-initiated withdrawal from the race).

Compared with other plain fences (no ditch), the first fence is almost seven times more likely to result in a fall (odds ratio, 6.8; 95% confidence interval, 4.6, 10.0). Becher's Brook (odds ratio, 4.7; 95% CI, 3.2, 7.1), open-ditch fences (odds ratio, 2.4; 95% CI, 1.6, 3.6), Canal Turn (odds ratio, 2.1; 95% CI,

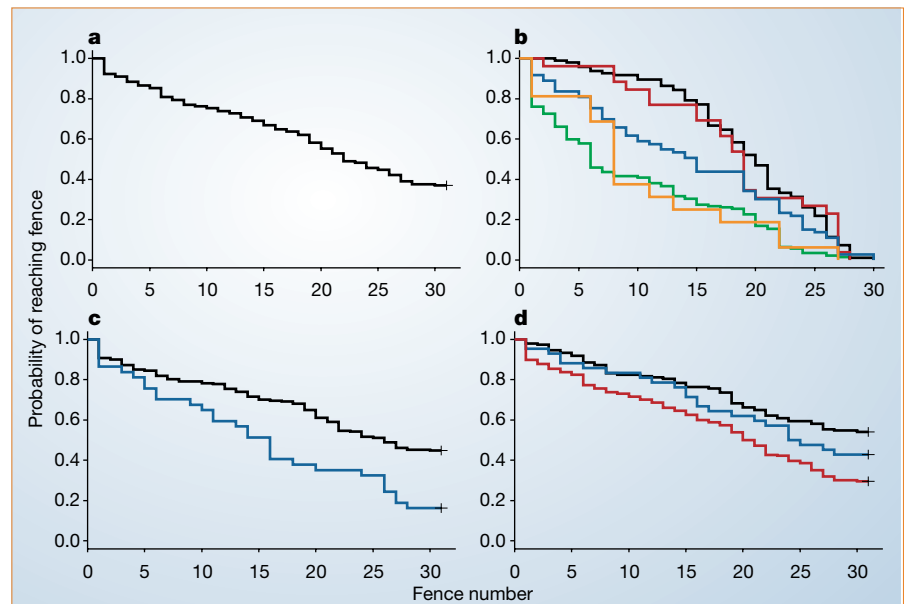


Figure 2 Patterns of non-completion of the Grand National during the past 15 races. **a**, Cumulative probability of reaching each fence for all 560 starters. **b**, Probability of reaching each fence is categorized by non-completion type: green, fell; black, pulled up; red, refused; blue, unseated; orange, brought down. **c**, Probability of reaching each fence categorized by ground condition ('going'): black, good or good-to-soft; blue, soft or heavy. **d**, Probability of reaching each fence categorized by previous experience on the Grand National course: black, raced, no falls; blue, raced and fell; red, never raced.

1.2, 3.6) and The Chair (odds ratio, 2.3; 95% CI, 1.1, 4.6) all confer a significantly increased risk of falling.

Ground classified as 'soft' or 'heavy' resulted in a significantly higher rate of non-completion (hazard ratio, 2.05; 95% CI, 1.32, 3.17) compared with all other types. The model predicts that only 18% of runners will complete when the 'going' is soft or heavy (Fig. 2c). Horses that had never raced on the course were twice as likely not to complete it (hazard ratio, 2.0; 95% CI, 1.52, 2.63) as horses that had had previous experience without a fall (Fig. 2d); this relationship remained after adjusting for age. The number of runners in the race was associated with non-completion (hazard ratio, 1.09 per runner; 95% CI, 1.01, 1.13), as was the

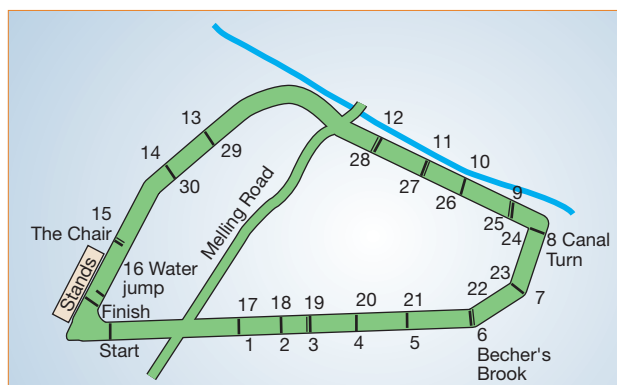
'starting-price multiplier' (the inverse of starting price; hazard ratio, 1.004; 95% CI, 1.002, 1.006) — the longer the odds, the greater the probability of non-completion.

Falling or unseating of the jockey as a result of a jumping error is the outcome most likely to result in injury²⁻⁴. Horses with no previous experience on the course were at increased risk of falling and of losing their rider (hazard ratio, 2.6; 95% CI, 1.80, 3.87). Good-to-soft ground was associated with a significantly decreased risk of falling relative to good ground (hazard ratio, 0.41; 95% CI, 0.21, 0.78). The number of runners in the race (hazard ratio, 1.08; 95% CI, 1.01, 1.17) and the starting-price multiplier (hazard ratio, 1.003; 95% CI, 1.001, 1.005) were also significantly associated with a risk of falling.

Our results indicate that increasing the schooling of horses over fences of the type used for the Grand National, or in qualifying races over the Grand National course, and providing good-to-soft ground would improve completion figures and decrease the risk of horses falling during the race. The greater risk of falling at the first fence indicates that interventions involving this obstacle could have a big impact on the number of horses that complete the race.

These findings also illustrate the value of using epidemiological methods to identify risk factors for injury in competitive sport and should inform the efforts of the racing

Figure 1 Layout of the course of the Grand National. The race comprises almost two complete circuits of this course. The 30 obstacles (numbered) jumped during the race vary in construction and are unique to the course. We classified fences into seven different categories: plain fences (no ditch), open ditch (ditch in front), first fence, water jump, Becher's Brook (ditch on landing side), Canal Turn (sharp turn immediately after a plain fence) and The Chair (high fence with a big ditch in front).



industry to safeguard equine welfare in the Grand National steeplechase.

Christopher Proudman, Gina Pinchbeck, Peter Clegg, Nigel French

Epidemiology Group, Faculty of Veterinary Science, University of Liverpool, Leahurst, Neston, Wirral CH64 7TE, UK

e-mail: c.j.proudman@liverpool.ac.uk

1. www.aintree.co.uk

2. McKee, S. L. *Eq. Vet. Educ.* **7**, 202–204 (1995).

3. Tuner, M., McCrory, P. & Halley, W. *Br. J. Sports Med.* **36**, 430–409 (2002).

4. Williams, R. B., Harkins, L. S., Hammond, C. J. & Wood, J. L. N. *Eq. Vet. J.* **33**, 478–486 (2001).

Competing financial interests: declared (see online brief communication for details).

Supplementary information accompanies this communication on Nature's website.

Microperiodic structures

Direct writing of three-dimensional webs

Applications are emerging that require the creation of fine-scale structures in three dimensions — examples include scaffolds for tissue engineering¹, microfluidic devices² and photonic materials that control light propagation over a range of frequencies³. But writing methods such as dip-pen nanolithography⁴ and ink-jet printing⁵ are either confined to two dimensions or beset by wetting and spreading problems. Here we use concentrated polyelectrolyte inks to write three-dimensional microperiodic structures directly without using masks. Our technique enables us to write arbitrary three-dimensional patterns whose features are nearly two orders of magnitude smaller than those attained with other multilayer printing techniques⁶.

For this direct-write process, we have developed fluid inks that flow readily through fine deposition nozzles and solidify rapidly in a coagulation reservoir (Fig. 1). The inks are made up of concentrated polyelectrolyte complexes that consist of non-stoichiometric mixtures⁷ of polyanions (polyacrylic acid, PAA) and polycations (polyethylenimine, PEI, or polyallylamine hydrochloride, PAH). By regulating the ratio of anionic ($\text{COO}^- \text{Na}^+$) to cationic (NH_3^+) groups and combining these species under solution conditions that promote polyelectrolyte exchange reactions⁸, we produce homogeneous fluids (40–50 wt % polyelectrolyte in aqueous solution; for methods, see supplementary information) with the requisite viscosity (about 5–150 pascal seconds) for deposition through microcapillary nozzles of varying diameter (0.5–5.0 μm).

When deposited in an alcohol/water coagulation reservoir, these concentrated polyelectrolyte inks coagulate to form self-supporting filaments or rods (Fig. 1). The reservoir composition strongly affects both the ink's elasticity and the coagulation mecha-

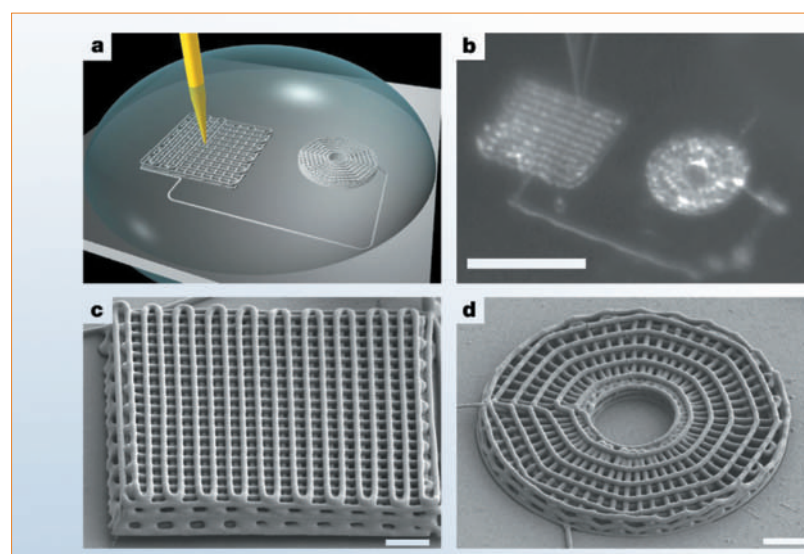


Figure 1 Direct-write assembly of three-dimensional microperiodic structures. **a**, The ink-deposition process (not drawn to scale). A concentrated polyelectrolyte ink is housed in a syringe (yellow) immersed in a coagulation reservoir (grey hemispherical drop) and deposited on to a glass substrate (light grey). **b**, Optical image acquired *in situ* during deposition reveals the features drawn in **a**, including the deposition nozzle that is patterning a three-dimensional lattice, as well as a completed radial array. This image is blurred by the reservoir (scale bar: 100 μm). **c**, Three-dimensional periodic structure with a face-centred tetragonal geometry (filament diameter: 1 μm ; 10 layers; scale bar: 10 μm). **d**, Three-dimensional radial array (filament diameter: 1 μm ; 5 layers; scale bar: 10 μm).

nism, which is driven by electrostatic effects in water-rich reservoirs and by solvent effects in alcohol-rich reservoirs. For example, the elasticity of PAA/PEI ink (ratio of anionic/cationic groups: $[\text{COO}^- \text{Na}^+]/[\text{NH}_3^+] = 5.7$) rises dramatically from 1 Pa (fluid phase) to nearly 10^5 Pa (coagulated phase) upon deposition in a reservoir containing 83–88% isopropyl alcohol. Under these conditions, the deposited ink filament is elastic enough to retain its shape while being able to flow and adhere to the substrate and underlying patterned layers.

Images of three-dimensional microperiodic lattices and radial arrays assembled from the PAA/PEI ink are shown in Fig. 1c,d. These structures can have solid or porous walls and rod-like filaments that span the web, as well as tight and broad-angled features. A three-axis (x,y,z) micropositioning device deposits the ink, sequentially building layered, patterned structures (see supplementary information). After producing a two-dimensional patterned layer, the nozzle is raised in the z -direction to generate the next layer. The process is repeated until the full three-dimensional structure is fabricated (building time is about 5 min).

The versatility of these inks is demonstrated by microperiodic structures made from PAA/PAH ink ($[\text{COO}^- \text{Na}^+]/[\text{NH}_3^+] = 0.5$), which consist of a patterned array of polyelectrolyte filaments with a net positive surface charge (see supplementary information). These three-dimensional polyelectrolyte scaffolds open up new opportunities for the electrostatic layer-by-layer assembly⁹ of materials, which has previously been limited to thin-film⁹ or discrete colloidal structures¹⁰.

Direct-write assembly is a powerful way to create three-dimensional microscale structures of arbitrary design and functionality. A

variety of inks could be developed using other polyelectrolyte mixtures, based for example on biologically, electrically or optically active polyelectrolytes. The structures created could direct cell-scaffold interactions¹¹, manipulate fluid flow², control light propagation³, or respond to environmental stimuli¹².

Gregory M. Gratson[†], Mingjie Xu[†], Jennifer A. Lewis^{†*}

**Frederick Seitz Materials Research Laboratory and Materials Science and Engineering Department, and [†]Chemical and Biomolecular Engineering Department, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA e-mail: jalewis@uiuc.edu*

1. Griffith, L. G. & Naughton, G. *Science* **295**, 1009–1014 (2002).

2. Theriault, D., White, S. R. & Lewis, J. A. *Nature Mater.* **2**, 265–271 (2003).

3. Lin, S. Y. *et al. Nature* **394**, 251–253 (1998).

4. Piner, R. D., Zhu, J., Xu, F., Hong, S. & Mirkin, C. A. *Science* **283**, 661–663 (1999).

5. Sirringhaus, H. *et al. Science* **290**, 2123–2126 (2000).

6. Chrisey, D. B. *Science* **289**, 879–881 (2000).

7. Zevin, A. B. & Kabanov, V. A. *Russ. Chem. Rev.* **51**, 833–855 (1982).

8. Zintchenko, A., Rother, G. & Dautzenberg, H. *Langmuir* **19**, 2507–2513 (2003).

9. Decher, G. *Science* **277**, 1232–1237 (1997).

10. Caruso, F., Caruso, R. A. & Möhwald, H. *Science* **282**, 1111–1114 (1998).

11. Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. *Science* **276**, 1425–1428 (1997).

12. Lee, Y.-J. & Braun, P. V. *Adv. Mater.* **15**, 563–566 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

brief communications arising online

www.nature.com/bca

Cosmology: Synchrotron radiation and quantum gravity

J. Ellis, N. E. Mavromatos, D. V. Nanopoulos, A. S. Sakharov (doi:10.1038/nature02481)

Cosmology

Synchrotron radiation and quantum gravity

Arising from: Jacobson, T., Liberati, S. & Mattingly, D. *Nature* **424**, 1019–1021 (2003)

Quantum gravity may cause the vacuum to act as a non-trivial medium (space–time foam), which alters the standard Lorentz relation between the energy and momentum of matter particles, thereby modifying their dispersion relations. Jacobson, Liberati and Mattingly¹ argue that synchrotron radiation from the Crab nebula imposes a stringent constraint on any modification of the dispersion relation of the electron that might be induced by quantum gravity, but their analysis does not constrain any modification of the dispersion relation of the photon^{2,3}. Such quantum-gravity effects need not obey the equivalence principle⁴ in the sense of being universal for all matter particles, as exemplified by quantum-gravity models in which photons are the only standard-model particles able to ‘see’ special quantum-gravity configurations that modify their dispersion relations. This implies that photons may be the only sensitive probe of quantum-gravity effects on particle dispersion relations, and the results of Jacobson *et al.* do not exclude all possible modifications of dispersion relations, even if they are suppressed by only a single power of the Planck mass (the characteristic quantum-gravity scale) — contrary to some subsequent interpretations of their results.

As pointed out previously⁴, there are theoretical models in which quantum gravity produces Lorentz invariance-violating effects for neutral particles, such as the photon, but not charged particles, such as the electron. One model of space–time foam³ suggests a linear modification of the dispersion relation for the photon: $p_\gamma = E_\gamma - (E_\gamma^2/M_{\text{QG}})$, where p_γ (E_γ) is the photon’s momentum (energy) and M_{QG} is some characteristic scale associated with quantum gravity, which may be of the same order as the Planck mass $M_{\text{P}} \approx 10^{19}$ GeV. However, this model³ predicts that there is no such modification of the dispersion relation for the electron⁴, and hence is compatible with the constraint¹ from the Crab nebula. In such models, constraints on the electron and nucleon dispersion relations^{1,5,6} are irrelevant,

leaving measurements on time profiles of very remote γ -ray bursts^{2,7} as the best approach for probing quantum-gravity effects.

The basic reason for this violation of the equivalence principle in the quantum-gravity model³ is its description of space–time foam by using quantum defects in space–time with vacuum quantum numbers, as in one interpretation of Liouville string theory⁸. These can be excited only by particles that are neutral under the gauge group of the standard model, such as photons, and such interactions give the vacuum a non-trivial refractive index for light of different frequencies (energies)². Charged particles, such as electrons, cannot form such excitations, so do not ‘see’ the space–time foam at all, and hence obey the usual Lorentz kinematics. As a result of the excitation of the vacuum by an energetic photon, space–time is distorted and the photon travels with a velocity smaller than the (supposedly universal) speed of light *in vacuo*, c , as postulated in the special and general theories of relativity.

As the electron has no interaction with the quantum-gravitational vacuum medium in this approach, it emits no Čerenkov radiation, despite travelling faster than photons, thus avoiding the vacuum Čerenkov radiation constraint⁹, as well as the Crab nebula constraint derived by Jacobson *et al.*¹ The model in ref. 3 also avoids the strong constraints described in ref. 10, as well as many other constraints on quantum-gravity effects¹¹. Claims that modified dispersion relations for photons would result in phase incoherence of light, and thereby destroy diffraction patterns in images of extragalactic sources¹², have been criticized by Ng¹³, who pointed out that the induced incoherent effects had been overestimated¹² by a large factor. In the specific model of ref. 3, the re-emission of the photon by a space–time defect is accompanied by a random phase in its wave function, destroying any cumulative phase incoherence. Finally, we note that, as the nucleon is a bound state, it is more complex to analyse, but we also do not expect it to exhibit a linear modification of the normal

Lorentz-invariant dispersion relation, avoiding other constraints^{5,6}.

The strong bound of Jacobson *et al.*¹ on the electron underlines the interest in probing directly the dispersion relation of the photon. The study of the arrival times of photons from γ -ray bursts² still appears to be the best experimental probe of any possible refractive index for photons, and should be pursued further. It has already established a lower limit on M_{QG} close to 10^{16} GeV (ref. 7), and current (HETE, INTEGRAL) and future (GLAST, AMS) high-energy space missions have the potential to reach the Planck scale for any linear quantum-gravity modification of the photon’s dispersion relation.

John Ellis*, N. E. Mavromatos†‡, D. V. Nanopoulos§¶, A. S. Sakharov**

*Department of Physics, CERN Theory Division, 1211 Geneva 23, Switzerland

e-mail: Alexandre.Sakharov@cern.ch

†Department of Physics, King’s College London, University of London, London WC2R 2LS, UK

‡Departamento de Física Teórica, Universidad de Valencia, 46100, Burjassot, Valencia, Spain

§George P. and Cynthia W. Mitchell Institute for Fundamental Physics, Texas A&M University, College Station, Texas 77843, USA

¶Astroparticle Physics Group, Houston Advanced Research Center, Mitchell Campus, Woodlands, Texas 77381, USA

¶Academy of Athens, Division of Natural Sciences, Athens 10679, Greece

#Swiss Institute of Technology, ETH–Zürich, 8093 Zürich, Switzerland

doi:10.1038/nature02481

1. Jacobson, T., Liberati, S. & Mattingly, D. *Nature* **424**, 1019–1021 (2003).
2. Amelino-Camelia, G., Ellis, J., Mavromatos, N. E., Nanopoulos, D. V. & Sarkar, S. *Nature* **393**, 763–765 (1998).
3. Ellis, J., Mavromatos, N. E. & Nanopoulos, D. V. *Phys. Rev. D* **62**, 084019(1–10) (2000).
4. Ellis, J., Mavromatos, N. E. & Sakharov, A. S. *Astropart. Phys.* **20**, 669–682 arXiv:astro-ph/0308403 (2004).
5. Carroll, S. *Nature* **424**, 1007–1008 (2003).
6. Myers, R. C. & Pospelov, M. *Phys. Rev. Lett.* **90**, 211601(1–4) (2003).
7. Ellis, J., Mavromatos, N. E., Nanopoulos, D. V. & Sakharov, A. S. *Astron. Astrophys.* **402**, 409–424 (2003).
8. Ellis, J., Mavromatos, N. E. & Nanopoulos, D. V. *J. Chaos Solit. Fract.* **10**, 345–362 (1999).
9. Liberati, S., Jacobson, T. & Mattingly, D. *Phys. Rev. D* **67**, 124011(1–26) (2003).
10. Alfaro, J. & Palma, G. *Phys. Rev. D* **67**, 083003(1–18) (2003).
11. Ellis, J., Mavromatos, N. E. & Nanopoulos, D. V. *Phys. Rev. D* **65**, 064007(1–7) (2002).
12. Lieu, R. & Hillman, L. W. *Astrophys. J.* **585**, L77–L80 (2003).
13. Ng, Y. J. *Mod. Phys. Lett. A* **18**, 1073–1097 (2003).

Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo

Masamichi Yamamoto¹, Yukio Saijoh¹, Aitana Perea-Gomez^{2*}, William Shawlot^{3*}, Richard R. Behringer³, Siew-Lan Ang^{2,4}, Hiroshi Hamada¹ & Chikara Meno¹

¹Developmental Genetics Group, Graduate School of Frontier Biosciences, Osaka University, and CREST, Japan Science and Technology Corporation (JST), 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan

²IGBMC/CNRS/INSERM, Université Louis Pasteur, 67404 Illkirch Cedex, CU de Strasbourg, France

³Department of Molecular Genetics, M.D. Anderson Cancer Center, University of Texas, Houston, Texas 77030, USA

⁴Division of Developmental Neurobiology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

* Present addresses: Institut Jacques Monod, CNRS, Universités Paris 6 et 7, 75251 Paris Cedex 05, France (A.P.-G.); Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, Minnesota 55455, USA (W.S.)

Patterning of the mouse embryo along the anteroposterior axis during body plan development requires migration of the distal visceral endoderm (DVE) towards the future anterior side by a mechanism that has remained unknown. Here we show that Nodal signalling and the regionalization of its antagonists are required for normal migration of the DVE. Whereas Nodal signalling provides the driving force for DVE migration by stimulating the proliferation of visceral endoderm cells, the antagonists *Lefty1* and *Cerl* determine the direction of migration by asymmetrically inhibiting Nodal activity on the future anterior side.

Establishment of the anteroposterior (A–P) axis is the first overt manifestation of the body plan in the mouse^{1,2}. The first morphological sign of A–P patterning is the formation of the primitive streak on the posterior side of the embryo at embryonic day (E)6.5. At the cellular and molecular levels, however, the patterning events begin much earlier. One day before gastrulation (E5.5), the proximodistal axis is established; at this time, the embryo appears radially symmetric, with a group of visceral endoderm (VE) cells that express the homeobox gene *Hex* being present exclusively at the distal end³. A few hours later, at E5.7, these *Hex*-positive distal visceral endoderm (DVE) cells begin to migrate towards the future anterior side, eventually forming the anterior visceral endoderm (AVE) at E6.5. The AVE secretes signalling molecules, including the Nodal antagonists *Lefty1* (ref. 4) and *Cerl* (refs 5, 6), both of which will inhibit Nodal signalling in the underlying epiblast and restrict Nodal activity to the posterior side of the embryo. Such asymmetry in Nodal signalling specifies that the anterior epiblast will form the prospective brain and generates the primitive streak on the posterior side.

Although the mechanism by which the AVE patterns the epiblast has been relatively well characterized^{7,8}, it has remained unknown how DVE migration is achieved and how the direction of DVE migration is determined. Here we address this important issue.

Expression of Nodal antagonists before DVE migration

We have previously shown that DVE migration does not occur in the most severe phenotype (type B) of embryos that lack the *FoxH1* transcription factor⁹. Examination of various types of *FoxH1* mutant embryos for *Hex* and *Nodal* expression revealed an apparent correlation between the level of *FoxH1*-mediated Nodal signalling in the VE and timing of the onset of DVE migration (Supplementary Figs S1 and S2). This suggested that Nodal signalling might promote DVE migration and that asymmetry in Nodal activity may determine the direction of such migration. However, *Nodal* expression is symmetric along the prospective A–P axis at E5.5 and E5.7 (ref. 10), the latter being the time when the DVE begins to migrate. We therefore examined in detail the expression domains of

the genes for the Nodal antagonists *Lefty1* and *Cerl* at the stages before and after DVE migration (between E5.0 and E6.5).

Expression of *Lefty1* and *Cerl* was first apparent at E5.5 in the DVE. Unexpectedly, their expression domains were already shifted towards one side of the embryo before the onset of DVE migration (Fig. 1a, d). From E6.0 to E6.5 (Fig. 1b, c, e, f), the expression domains of *Lefty1* and *Cerl* moved towards the anterior side. These observations suggested that the expression domains of *Lefty1* and *Cerl* at E5.5 are biased towards the prospective anterior side. This conclusion was confirmed by two-colour *in situ* hybridization. Both *Lefty1* (Fig. 1g) and *Cerl* (Fig. 1i) expression domains at E5.5 were shifted towards the future anterior side of embryos ($n = 8$), whereas the *Hex* expression domain (Fig. 1h, j) was apparent at the distal tip of the same embryos. Such asymmetric expression of *Lefty1* and *Cerl* might thus be expected to generate differential Nodal activity in the DVE along the A–P axis, with lower Nodal activity on the future anterior side.

Asymmetric Nodal inhibition directs DVE migration

To determine whether an asymmetry in Nodal signalling generated by the Nodal antagonists directs DVE migration, we examined the effects of ectopic expression of *Nodal*, *Lefty1* or *Cerl*, each together with the gene for green fluorescent protein (GFP), in the VE of E5.2 embryos. The DVE was labelled with the dye DiI³ to trace its migration. Embryos were cultured for 24 h and the positions of GFP-positive and DiI-positive cells were then mapped (Fig. 2a–c). If the test gene did not influence the direction of DVE migration, we would expect GFP-positive cells and DiI-positive cells to be found on random sides of the embryo; the probability of both types of cells being located on the same side would be 25%. In contrast, if the Nodal antagonists were able to direct DVE migration, GFP-positive cells and DiI-positive cells should be located on the same side of the embryo with a frequency of >25%.

Expression of GFP alone did not influence the direction of DVE movement; DiI-labelled cells were thus randomly situated relative to the position of GFP-positive cells, with a co-localization frequency of 26% ($n = 71$) (Fig. 2d). Ectopic expression of *Lefty1* resulted in

the co-localization of GFP-positive cells and DiI-positive cells on the same side in 54% of embryos ($n = 35$; $P < 0.005$ versus GFP alone, χ^2 test). Ectopic expression of *Cerl* had a similar effect on DVE migration, with 46% of embryos ($n = 57$, $P < 0.03$) showing same-sided localization of the two types of cells. In embryos ectopically expressing both *Lefty1* and *Cerl*, the DVE migrated towards the side containing the GFP-positive cells in 88% of embryos examined ($n = 57$, $P < 0.0001$). The VE cells overexpressing *Lefty1* and *Cerl* were negative for *Hex* expression (Supplementary Fig. 3b, c), suggesting that they did not become AVE.

We similarly introduced a *Nodal* expression vector into wild-type

embryos. DiI and GFP signals were located on the same side in only 4% of such embryos ($n = 85$, $P < 0.0001$) (Fig. 2d), indicating that the DVE migrated away from the side with a higher *Nodal* activity. The DVE fails to migrate in *FoxH1*^{-/-} embryos containing one wild-type *Nodal* allele and a *Nodal-lacZ* allele¹⁰. In such *FoxH1*^{-/-}, *Nodal*^{lacZ/+} mutant embryos, *Nodal* expression is lost in the VE and distal epiblast and, as in *FoxH1*^{-/-} type B embryos, that in the proximal epiblast is radially symmetric. If the migration failure in this mutant is due to the loss of *Nodal* expression, restoration of *Nodal* expression would be expected to induce DVE movement. Indeed, whereas DiI-labelled cells remained at the distal tip in 100%

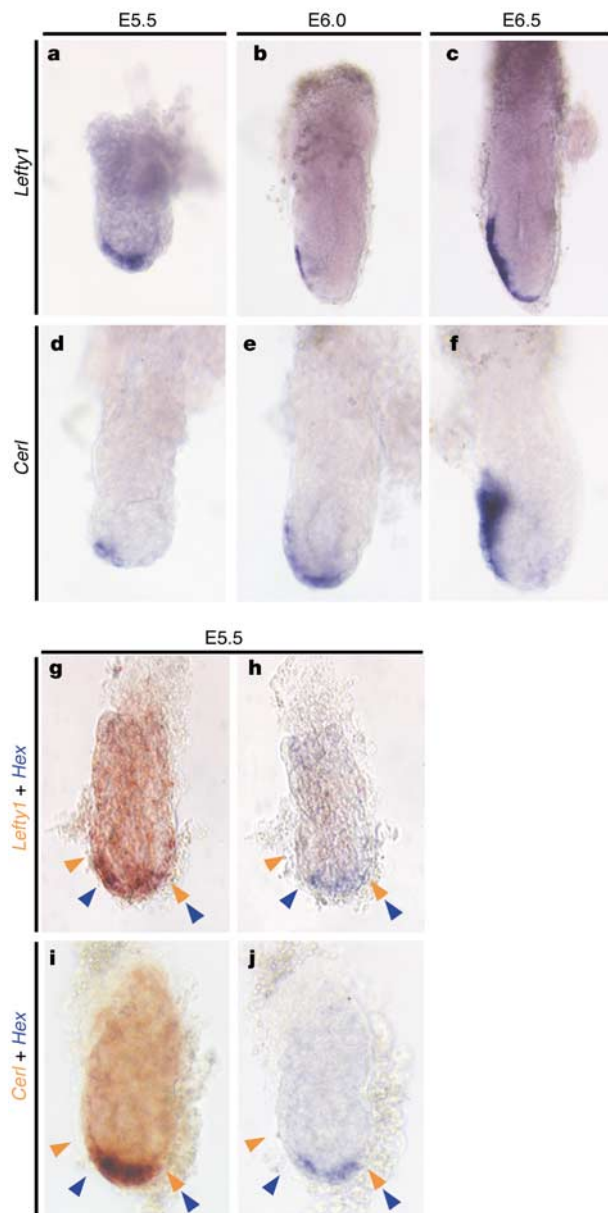


Figure 1 Asymmetric expression of *Lefty1* and *Cerl* before DVE migration. Expression of *Lefty1* (a–c) and *Cerl* (d–f) was examined in wild-type mouse embryos at the indicated stages of development. Expression of *Lefty1* (orange) and *Hex* (blue) (g) or of *Cerl* (orange) and *Hex* (blue) (i) was simultaneously examined in the same E5.5 embryos by two-colour *in situ* hybridization; the *Hex* expression domains alone (h, j) were visualized after removal of the orange staining by treatment of the embryos shown in g and i with methanol. Orange arrowheads indicate the borders of *Lefty1* (g, h) or *Cerl* (i, j) expression domains; blue arrowheads indicate those of *Hex* expression domains (g–j). Lateral views are shown for each embryo, with the anterior side on the left.

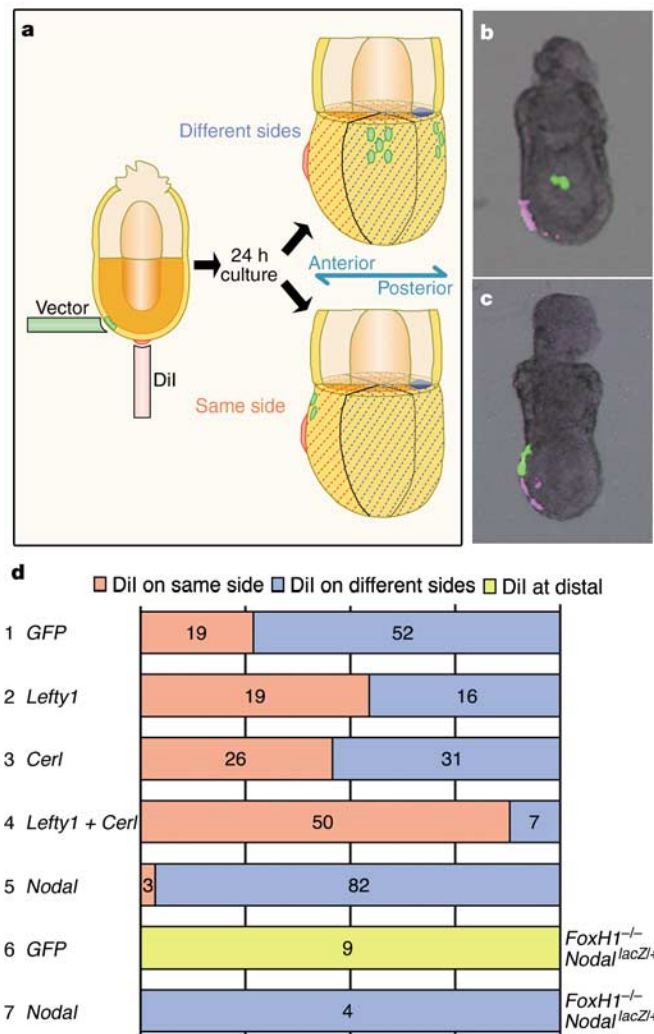


Figure 2 Ectopic expression of *Lefty1* or *Cerl* directs DVE migration. **a**, Experimental strategy. An effector gene, together with a GFP expression vector, was introduced into VE cells on the lateral side of E5.2 mouse embryos, and the DVE was labelled with DiI. The embryos were then cultured for 24 h, after which the fates of GFP-labelled cells (green) and DiI-labelled cells (red) were examined. The egg cylinder was divided into four quarters; localization of DiI-labelled cells in the same quarter or in a different quarter relative to the position of GFP-positive cells was categorized as 'DiI on same side' or 'DiI on different side', respectively. **b**, **c**, Two representative embryos showing localization of GFP-positive cells and DiI-positive cells on different sides (**b**) or on the same side (**c**). **d**, Summary of the effects of ectopic expression of the indicated genes on DVE migration. The numbers of embryos showing each localization pattern are indicated. Host embryos were wild-type with the exception of those in rows 6 and 7, which were *FoxH1*^{-/-}, *Nodal*^{lacZ/+}. 'DiI at distal' (yellow) indicates a pattern in which the DVE remains at the distal end without migration.

of *FoxH1*^{-/-}, *Nodal*^{lacZ/+} embryos expressing GFP alone ($n = 9$), ectopic expression of *Nodal* repelled the DVE in 100% of such embryos ($n = 4$, $P < 0.0001$ versus those expressing GFP alone) (Fig. 2d).

Nodal antagonists inhibit VE cell proliferation

To determine whether differential cell proliferation contributes to the mechanism of DVE migration, we examined the proliferation of VE cells by labelling embryos with bromodeoxyuridine (BrdU) at pre- or early gastrulation stages. In wild-type embryos, cell proliferation in the VE was uniform at E5.5 (Fig. 3a). At E5.7, however, when the DVE begins to migrate, cell proliferation was inhibited in the region corresponding to the *Lefty1* and *Cerl* expression domains (Fig. 3b, b'). Inhibition of cell proliferation in the future AVE was manifest at E6.0 (Fig. 3c, c'), and the asymmetry in cell proliferation rate along the A–P axis was maintained in the VE until E6.5 (Fig. 3d, e, e'). This pattern of cell proliferation was reminiscent of the temporal changes in Nodal signalling. To investigate the possible relation between Nodal signalling and cell proliferation, we examined the pattern of BrdU incorporation at E6.5 in various mutants deficient in such signalling. In *Nodal*^{lacZ/lacZ} embryos¹¹ ($n = 5$), cell proliferation was markedly impaired in the VE and epiblast (Fig. 3f). In *FoxH1*^{-/-} embryos, two types of staining pattern were observed. In the type A mutant ($n = 3$), an asymmetric pattern of BrdU incorporation similar to that apparent in the wild type was detected (Fig. 3g, g'), although the staining level was reduced compared with that in wild-type embryos. In the type B mutant ($n = 4$), which expresses *Nodal* only in the proximal epiblast at this stage (Supplementary Fig. S2; data not shown), uniform staining was apparent in the VE overlying the proximal epiblast whereas staining in the DVE was lost (Fig. 3h, h'). Finally, in *FoxH1*^{flax/-}, *Lefty2*-cre embryos ($n = 4$) lacking FoxH1 specifically in the epiblast⁹, BrdU incorporation in the proximal region was similar to that apparent in wild-type embryos (Fig. 3i, i'). However, the anterior DVE was more strongly stained than that of the wild-type embryo (Fig. 3i), because these mutant embryos lack Lefty1 in the AVE⁹. These results indicate that Nodal signalling promotes cell proliferation in the VE and that the asymmetry in proliferation

along the A–P axis is due to differential Nodal signalling.

We next examined the effects of ectopic expression of Nodal or Nodal antagonists on the proliferation of VE cells (Fig. 4). The VE of E5.2 embryos was co-transfected unilaterally with a test gene and a GFP expression vector. The DVE was labelled with DiI in some embryos (37/86 for GFP alone; 72/101 for Nodal; 32/86 for Lefty1 + Cerl). The number of GFP⁺ cells was counted 12 and 36 h later, and the increase in the number of GFP⁺ cells during the 24-h period between these two time points (referred to as replication index, RI) was estimated. In embryos transfected with the GFP vector alone ($n = 86$), the RI of the GFP⁺ VE cells varied between 1.0 and 9.0 (Fig. 4a–d, i). This variability depends on the specific site of transfection, with transfected VE cells on the future anterior side dividing fewer times (average RI = 2.2, $n = 13$) than those on other sides (right side, average RI = 4.0, $n = 9$; left side, average RI = 3.7, $n = 8$; posterior side, average RI = 4.2, $n = 7$). Transfection of embryos with a *Nodal* expression vector ($n = 101$) stimulated cell proliferation (Fig. 4g–i), as reflected by a marked increase in the percentage of embryos in which cells replicated more than four times. In most of the embryos with the DVE labelled with DiI (69/72), the GFP⁺ cells were located on the non-anterior sides with an average RI of 7.2, which was markedly higher than the RI values of control VE cells: 4.0, 3.7, 4.2 and 2.2. Conversely, ectopic expression of both *Lefty1* and *Cerl* in embryos ($n = 86$) resulted in inhibition of cell proliferation (Fig. 4e, f, i), with transfected cells rarely duplicating more than four times. In most of the embryos with the DVE labelled with DiI (23/32), GFP⁺ cells were found on the anterior side with an average RI of 2.5, which is significantly lower than that of the control VE cells. These results demonstrate that the DVE migrates away from the side where VE proliferation is stimulated by ectopic expression of Nodal. Conversely, the DVE migrates towards the side where proliferation of VE cells is inhibited by Lefty1 and Cerl.

To determine the extent to which asymmetric cell proliferation contributes to DVE migration, we subjected embryos both to unilateral co-transfection of the VE with a GFP vector and an expression vector for wild-type or a dominant negative form of *Cdk2* (ref. 12) as well as to labelling of the DVE with DiI (Fig. 4j).

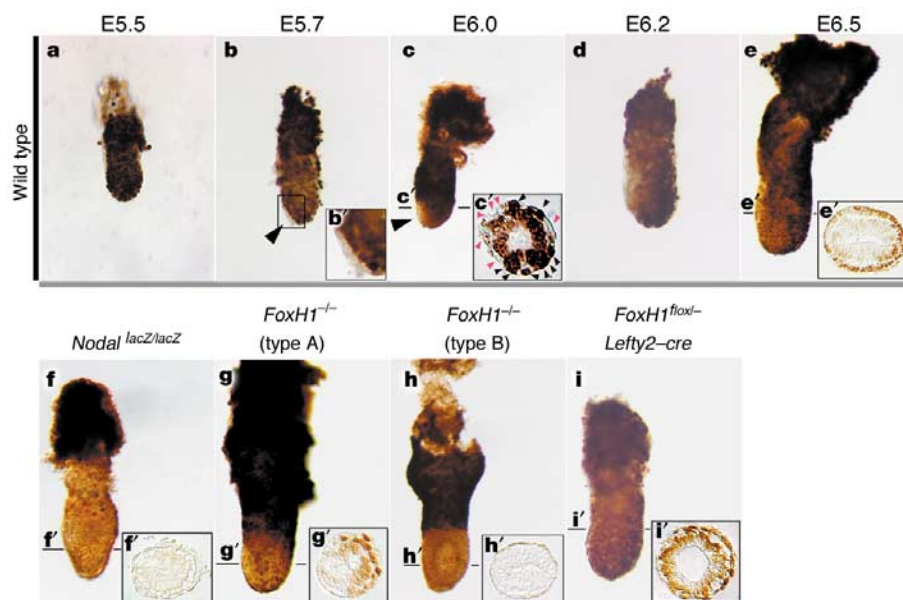


Figure 3 Patterns of VE cell proliferation in wild-type and various mutant embryos. The pattern of BrdU incorporation was determined in wild-type embryos at the indicated stages (top row: a–e) as well as in the indicated mutants at E6.5 (bottom row: f–i). Lateral views are shown for each embryo, with the anterior side on the left. Arrowheads indicate regions negative for BrdU labelling. The square region in b is expanded in b'. The

planes of transverse sections at the indicated points on each embryo (c', e', f', g', h' and i') are indicated by the horizontal bar of the embryos in c and e–i, respectively. c', BrdU-positive or -negative nuclei in the VE are shown by closed black and red arrowheads, respectively.

Ectopic expression of wild-type *Cdk2* repelled the DVE in 90% of embryos examined ($n = 79$, $P < 0.008$ versus embryos expressing GFP alone), whereas that of the dominant negative form of *Cdk2* attracted the DVE in 45% of embryos ($n = 98$, $P < 0.016$). We next examined the effect of ectopic expression of *Cdk2* in *FoxH1*^{-/-}, *Nodal*^{lacZ/+} embryos. Whereas expression of GFP alone failed to move the DVE, that of wild-type *Cdk2* repelled the DVE in all of the four embryos examined ($P < 0.0001$). These data suggest that inhibition of cell proliferation by Nodal antagonists directs the DVE to migrate towards the future anterior side.

Impaired anteroposterior patterning without Lefty1 and Cerl

To confirm the role of Nodal antagonists in A-P patterning, we analysed mutant mice lacking Lefty1 or Cerl. *Lefty1*^{-/-} (ref. 13) and *Cerl*^{-/-} (refs 6, 14, 15) embryos undergo normal gastrulation, whereas *Cerl*^{-/-}, *Lefty1*^{-/-} embryos develop multiple primitive streaks⁸, suggestive of functional redundancy between *Lefty1* and *Cerl*. We first examined expression of the AVE marker *Hex* (Fig. 5a, b, h, i). The *Hex* expression domain was expanded slightly in *Cerl*^{-/-} and *Lefty1*^{-/-} mutants (data not shown) and markedly in the *Cerl*^{-/-}, *Lefty1*^{-/-} double mutant (Fig. 5b, i), in which cells positive for *Hex* transcripts occupied the entire anterior half of the VE at E6.5 ($n = 4$). These results reveal that the Nodal antagonists restrict the size or location of the AVE, perhaps by inhibiting the

proliferation of AVE cells.

We next examined DVE migration in the *Lefty1* and *Cerl* mutants by monitoring *Hex* expression. Although the *Lefty1*^{-/-} or *Cerl*^{-/-} embryos did not manifest obvious migration defects (data not shown), migration of the DVE was delayed in *Cerl*^{-/-}, *Lefty1*^{-/-} embryos (Fig. 5o). The *Hex* expression domain, which is normally located on the anterior side at E6.0 (Fig. 5a), thus remained in the distal region of the double mutant at this time (Fig. 5h).

The asymmetry in BrdU incorporation between the anterior and posterior VE was maintained in *Lefty1*^{-/-} ($n = 4$) (Fig. 5s, s') and *Cerl*^{-/-} ($n = 5$) (Fig. 5w, w') embryos. In *Cerl*^{-/-}, *Lefty1*^{-/-} embryos at E6.0 ($n = 5$) (Fig. 5j) and E6.5 ($n = 4$) (Fig. 5k, k'), BrdU incorporation was apparent at a high level throughout the entire VE, but a small BrdU-negative region remained in the AVE. We also examined *Nodal* expression in the various mutant embryos at E6.5. In the wild type, *Nodal* expression in the epiblast at this stage is greater in the posterior region² (Fig. 5e–g). In the single mutants lacking *Lefty1* ($n = 3$) (Fig. 5t–v) or *Cerl* ($n = 5$) (Fig. 5x–z), *Nodal* expression was upregulated in the VE. Increased *Nodal* expression was more evident in the double mutant ($n = 4$) (Fig. 5l–n); *Nodal* expression in the epiblast was thus uniform in the distal region (Fig. 5n) but was asymmetric in the proximal region (Fig. 5m). In addition, *Nodal* expression was upregulated throughout the VE (Fig. 5l–n).

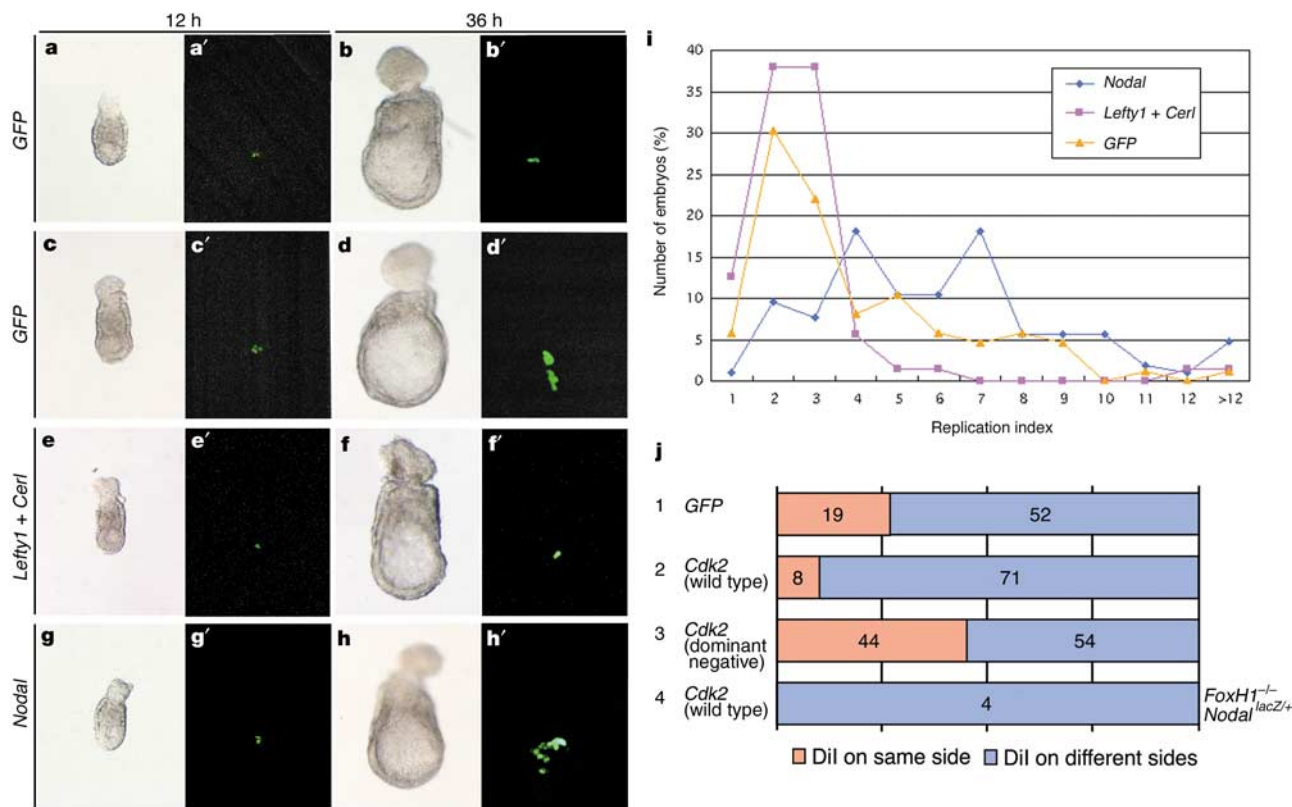


Figure 4 Nodal antagonists inhibit and Nodal promotes VE cell proliferation. **a–h**, The VE of E5.2 wild-type embryos was transfected unilaterally with a GFP expression vector either alone or together with an effector gene, as indicated. The DVE was labelled with Dil in some embryos (37/86 for GFP alone; 72/101 for Nodal; 32/86 for Lefty1 + Cerl). The numbers of GFP-positive cells were counted 12 and 36 h after transfection. Both differential interference contrast (**a–h**) and GFP fluorescence (**a'–h'**) images of typical embryos are shown for the three groups. Two types of embryos are shown for those that received GFP alone. The embryo shown in **a**, **a'**, **b** and **b'** represents the case where GFP was introduced into the prospective anterior side, whereas the one shown in **c**, **c'**, **d** and **d'** represents the case where GFP was introduced into non-anterior sides. **i**, Summary of the percentages of embryos in which GFP-positive cells replicated the indicated numbers

of times during the 24-h period of analysis for the three groups of embryos. The data are based on embryos in which the DVE was labelled with Dil and those unlabelled. Thus, the sample sizes are 86 for GFP alone, 101 for Nodal, and 86 for Lefty1 + Cerl. **j**, Effects of ectopic expression of wild-type *Cdk2* or of a dominant negative form of *Cdk2* on DVE migration. Host embryos were wild-type (rows 1–3) or *FoxH1*^{-/-}, *Nodal*^{lacZ/+} (row 4) and were analysed as described in Fig. 2a. DVE migration induced by *Cdk2* in *FoxH1*^{-/-}, *Nodal*^{lacZ/+} embryos was incomplete (the migration distance was shorter than that in wild-type embryos), possibly because the small population of proliferating cells was not sufficient for complete migration of the DVE or because Nodal signalling has relevant effects other than on cell proliferation. The numbers of embryos showing localization of Dil on the same side or on a different side relative to that of GFP are indicated.

Although the extent of VE cell proliferation was greatly increased in the double mutant, a small BrdU-negative region was apparent in the AVE (Fig. 5j, k, k'). These observations suggest the existence of an additional Nodal antagonist (or antagonists) that partially compensates for the lack of *Lefty1* and *Cerl*. An obvious candidate for such a molecule was *Lefty2*. *Lefty2* is not expressed in the VE of wild-type embryos⁴, but, unexpectedly, it was expressed in a portion of the AVE of *Cerl*^{-/-}, *Lefty1*^{-/-} embryos at E5.7 (*n* = 2) (Fig. 5p), E6.2 (*n* = 2) (Fig. 5q)¹⁶ and E6.5 (*n* = 5) (Fig. 5p) that corresponded to the BrdU-negative region (Fig. 5j, k). It is therefore likely that ectopic expression of *Lefty2* partially compensates for the loss of *Lefty1* and *Cerl*.

Discussion

How does differential Nodal activity drive DVE migration? Several

lines of evidence indicate that differential cell proliferation in the VE along the future A-P axis regulates DVE migration (Supplementary Fig. S4). However, they do not necessarily exclude other possibilities. In particular, migration of AVE, which is a fast cell migration event, may also involve an active migration mechanism. Such a mechanism should also require Nodal signal, because DVE fails to move in various mutant embryos lacking Nodal signal. It is interesting to note that DVE of the *FoxH1*^{-/-}, *Nodal*^{lacZ/+} embryos moved partially in response to Cdk2 (Fig. 4j). It may be that differential cell proliferation determines the direction of DVE migration, while a driving force of AVE migration is provided mainly by an active migration mechanism and partially by unequal cell proliferation. In this regard, nearly normal formation of AVE in *Cerl*^{-/-}, *Lefty1*^{-/-} embryos may not necessarily be due to the compensatory activity of *Lefty2* but could also be due to active

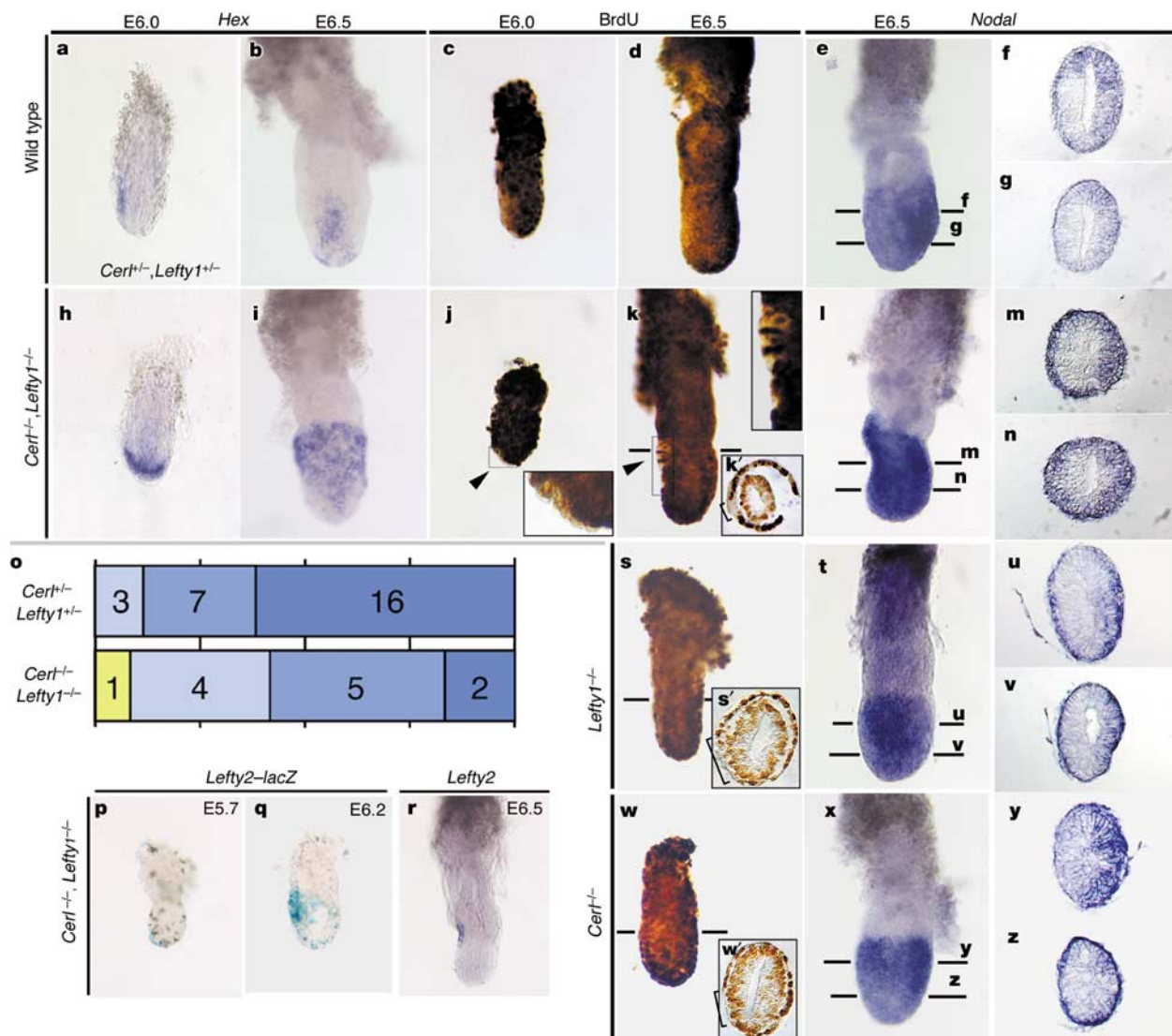


Figure 5 Anteroposterior patterning defects in the absence of *Lefty1* and *Cerl*. *Cerl*^{+/-}, *Lefty1*^{+/-} (a), wild-type (b–g), *Cerl*^{-/-}, *Lefty1*^{-/-} (h–n, p–r), *Lefty1*^{-/-} (s–v) and *Cerl*^{-/-} (w–z) embryos were examined for *Hex* expression (a, b, h, i), BrdU incorporation (c, d, j, k, s, w), *Nodal* expression (e–g, l–n, t–v, x–z), and *Lefty2* expression (r) at the indicated stages. *Lefty2* expression was also examined with a *Lefty2-lacZ* transgene (p, q) in which *LacZ* expression is driven by the upstream 5.5-kb region of the *Lefty2* gene¹⁷. This transgene is not active in the wild-type embryo at the pre-gastrulation stage (data not shown), but is expressed in the regions of the *Cerl*^{-/-}, *Lefty1*^{-/-} embryo at E5.7 (p) and E6.2 (q) that correspond to the BrdU-negative regions. Lateral views are shown for each

embryo with the anterior side on the left, with the exception that anterior views are shown in b and i. The planes of transverse sections (f, g, k', m, n, s', u, v, w', y, z) are indicated by the horizontal bars in e, k, l, s, t, w and x. Patterns of the *Hex* expression domain are summarized for *Cerl*^{+/-}, *Lefty1*^{+/-} and *Cerl*^{-/-}, *Lefty1*^{-/-} embryos at E6.0 in o. The numbers of embryos showing each pattern are indicated. Yellow, the expression domain remains at the distal tip; light blue, it was shifted towards the anterior side; medium blue, it reached half way towards the junction between embryonic and extraembryonic regions; dark blue, it reached the junction. BrdU-negative VE regions are indicated by arrowheads and brackets (j, k, k', s', w').

and non-Nodal-dependent cell movement. It is also possible that unequal cell proliferation determines the direction of migration indirectly. Thus, reduced proliferation may force cells to exit the cell cycle, which may induce the directional movement of neighbouring cells.

The ectopic expression of both *Lefty1* and *Cerl* had a greater effect on DVE migration than did the expression of either Nodal antagonist alone. Furthermore, A–P patterning defects were apparent only in the absence of both *Lefty1* and *Cerl*, although VE cell proliferation was mildly affected in *Lefty1*^{−/−} and *Cerl*^{−/−} single mutants. *Lefty1* and *Cerl* are thought to antagonize Nodal signalling via different mechanisms. Whereas *Lefty1* blocks the Nodal signal through competitive binding to the Nodal receptor¹⁷, *Cerl* inhibits Nodal activity by binding to Nodal itself¹⁸. Inhibition at both levels might thus be necessary to suppress Nodal signalling fully. Alternatively, *Cerl* may inhibit signalling by other molecules, such as Wnt or bone morphogenetic protein, as has been suggested previously^{6,18}. Mice deficient in β -catenin also show defects in DVE migration¹⁹, suggesting that a canonical pathway of Wnt signalling contributes to this process, either directly or indirectly.

The necessity of establishing polarity of *Nodal* expression along the A–P axis is conserved among vertebrates, although different strategies appear to be adopted to achieve this end. In *Xenopus* and zebrafish^{20,21}, the expression of Nodal-related proteins in the vegetal hemisphere or marginal zone is induced by maternal factors; this expression is later extinguished by a negative signalling loop involving *Lefty* (*Antivin*) in the zebrafish²², whereas the expression on the dorsal side is increased by β -catenin signalling²⁰ in addition to the Nodal positive signalling loop^{23,24}. In the chick, *Nodal* expression is induced by *Vg1* and Wnt in the posterior marginal zone²⁵, whereas the expression in the periphery of the blastoderm is suppressed by *Cerl* produced in the hypoblast²⁶. A positional cue such as that generated by cortical rotation in *Xenopus*²⁷ or by gravity in the chick²⁸ does not appear to contribute to mouse development. Instead, the mouse embryo may have developed a dynamic system involving *Nodal*–*Lefty*/*Cerl* positive and negative signalling loops that make use of a subtle change generated in the VE. □

Methods

Introduction of expression vectors into the VE

E5.2 embryos were dissected from the uterus and the parietal endoderm membrane was removed. The introduction of expression vectors (a GFP expression either alone or with various effectors)²⁹ into the VE and labelling of the DVE with *Dil* were performed with a Leica micromanipulator. Liposomes composed of the expression vectors and Lipofectamine2000 (Invitrogen) were injected between the VE and epiblast with an injection pipette (see Supplementary Information). Embryos were cultured for 24 h under a humidified atmosphere of 5% CO₂ at 37 °C in dishes containing Dulbecco's modified Eagle's medium supplemented with 75% rat serum. They were then examined with a Leica compound fluorescence microscope equipped with rhodamine and GFP2 optics³. In most injected embryos, the GFP signal was detected only in the VE; those in which GFP was apparent in the epiblast were discarded. However, we cannot exclude the possibility that an exogenous gene is expressed at a low level in the epiblast of some embryos.

BrdU labelling of embryos

Labelling of proliferating cells with BrdU was performed as described previously³⁰. In brief, pregnant mice were injected intraperitoneally with BrdU (100 mg per kilogram of body mass) at E5.5, E5.7, E6.0, E6.2 or E6.5 and killed 25 min thereafter. The embryos were recovered, fixed with Bouin's solution, and exposed consecutively to antibodies to BrdU (MBL, Japan), biotinylated secondary antibodies and an ABC-horseradish peroxidase system (Vector Laboratories). The genotype of each embryo was determined by polymerase chain reaction (PCR) with DNA obtained from the ectoplacental cone.

Two-colour whole-mount *in situ* hybridization

Mouse embryos were staged on the basis of their morphology. Two-colour whole-mount *in situ* hybridization was performed with one RNA probe labelled with digoxigenin and the other labelled with fluorescein according to a standard procedure. After the first staining, alkaline phosphatase was inactivated by incubation at 70 °C for 30 min. The second staining was performed with INT/BCIP solution (Roche), because the reaction product can later be decolorized by exposure to methanol.

Received 14 November 2003; accepted 16 February 2004; doi:10.1038/nature02418.

Published online 7 March 2004.

1. Beddington, R. S. & Robertson, E. J. Anterior patterning in mouse. *Trends Genet.* **14**, 277–284 (1998).

2. Beddington, R. S. & Robertson, E. J. Axis development and early asymmetry in mammals. *Cell* **96**, 195–209 (1999).
3. Thomas, P. Q., Brown, A. & Beddington, R. S. Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125**, 85–94 (1998).
4. Meno, C. *et al.* Two closely-related left-right asymmetrically expressed genes, *lefty-1* and *lefty-2*: their distinct expression domains, chromosomal linkage and direct neuralizing activity in *Xenopus* embryos. *Genes Cells* **2**, 513–524 (1997).
5. Belo, J. A. *et al.* Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mech. Dev.* **68**, 45–57 (1997).
6. Belo, J. A. *et al.* Cerberus-like is a secreted BMP and nodal antagonist not essential for mouse development. *Genesis* **26**, 265–270 (2000).
7. Kimura, C. *et al.* Visceral endoderm mediates forebrain development by suppressing posteriorizing signals. *Dev. Biol.* **225**, 304–321 (2000).
8. Perea-Gomez, A. *et al.* Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Dev. Cell* **3**, 745–756 (2002).
9. Yamamoto, M. *et al.* The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes Dev.* **15**, 1242–1256 (2001).
10. Norris, D. P. & Robertson, E. J. Asymmetric and node-specific nodal expression patterns are controlled by two distinct cis-acting regulatory elements. *Genes Dev.* **13**, 1575–1588 (1999).
11. Collignon, J., Varlet, I. & Robertson, E. J. Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* **381**, 155–158 (1996).
12. Belachew, S. *et al.* Cyclin-dependent kinase-2 controls oligodendrocyte progenitor cell cycle progression and is downregulated in adult oligodendrocyte progenitors. *J. Neurosci.* **22**, 8553–8562 (2002).
13. Meno, C. *et al.* *lefty-1* is required for left-right determination as a regulator of *lefty-2* and nodal. *Cell* **94**, 287–297 (1998).
14. Simpson, E. H. *et al.* The mouse *Cerl* (Cerberus related or homologue) gene is not required for anterior pattern formation. *Dev. Biol.* **213**, 202–206 (1999).
15. Shawlot, W., Min Deng, J., Wakamiya, M. & Behringer, R. R. The cerberus-related gene, *Cerr1*, is not essential for mouse head formation. *Genesis* **26**, 253–258 (2000).
16. Saijoh, Y. *et al.* Distinct transcriptional regulatory mechanisms underlie left-right asymmetric expression of *lefty-1* and *lefty-2*. *Genes Dev.* **13**, 259–269 (1999).
17. Sakuma, R. *et al.* Inhibition of Nodal signalling by *Lefty* mediated through interaction with common receptors and efficient diffusion. *Genes Cells* **7**, 401–412 (2002).
18. Piccolo, S. *et al.* The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature* **397**, 707–710 (1999).
19. Huelsen, J. *et al.* Requirement for beta-catenin in anterior-posterior axis formation in mice. *J. Cell Biol.* **148**, 567–578 (2000).
20. De Robertis, E. M., Larrain, J., Oelgeschlager, M. & Wessely, O. The establishment of Spemann's organizer and patterning of the vertebrate embryo. *Nature Rev. Genet.* **1**, 171–181 (2000).
21. Schier, A. F. Axis formation and patterning in zebrafish. *Curr. Opin. Genet. Dev.* **11**, 393–404 (2001).
22. Chen, Y. & Schier, A. F. *Lefty* proteins are long-range inhibitors of *Wnt*-mediated nodal signaling. *Curr. Biol.* **12**, 2124–2128 (2002).
23. Cheng, A. M., Thisse, B., Thisse, C. & Wright, C. V. The *lefty*-related factor *Xatv* acts as a feedback inhibitor of nodal signaling in mesoderm induction and L–R axis development in *Xenopus*. *Development* **127**, 1049–1061 (2000).
24. Meno, C. *et al.* Mouse *Lefty2* and zebrafish *antivin* are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4**, 287–298 (1999).
25. Skromme, I. & Stern, C. D. Interactions between Wnt and *Vg1* signalling pathways initiate primitive streak formation in the chick embryo. *Development* **128**, 2915–2927 (2001).
26. Bertocchini, F. & Stern, C. D. The hypoblast of the chick embryo positions the primitive streak by antagonizing nodal signaling. *Dev. Cell* **3**, 735–744 (2002).
27. Miller, J. R. *et al.* Establishment of the dorsal-ventral axis in *Xenopus* embryos coincides with the dorsal enrichment of *dishevelled* that is dependent on cortical rotation. *J. Cell Biol.* **146**, 427–437 (1999).
28. Eyal-Giladi, H., Goldberg, M., Refael, H. & Avner, O. A direct approach to the study of the effect of gravity on axis formation in birds. *Adv. Space Res.* **14**, 271–279 (1994).
29. Mizushima, S. & Nagata, S. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322 (1990).
30. Hayashi, Y., Koike, M., Matsutani, M. & Hoshino, T. Effects of fixation time and enzymatic digestion on immunohistochemical demonstration of bromodeoxyuridine in formalin-fixed, paraffin-embedded tissue. *J. Histochem. Cytochem.* **36**, 511–514 (1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Robertson for *Nodal*^{lacZ} mutant mice; R. Beddington for the *Hex* probe; V. Gallo for expression vectors for Cdk and dominant negative Cdk; and N. Mine and Y. Ohnishi for help in improving the lipofection procedure. We also thank S. Srinivas for exchanging unpublished data. This work was supported by grants from the Ministry of Education, Science, Sports, and Culture of Japan (to C.M. and H.H.), a grant from CREST (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation (to H.H.), a grant from the NIH (to R.R.B.), and a grant from the Association pour la Recherche sur le Cancer and by funds from the INSERM, the CNRS and the Hôpital Universitaire de Strasbourg (to S.-L.A.). M.Y. is a recipient of a fellowship from the Japan Society for the Promotion of Science for Japanese Junior Scientists.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.H. (hamada@fbs.osaka-u.ac.jp).

Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons

Jinsong Li, Tomohiro Ishii, Paul Feinstein & Peter Mombaerts

The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Of the ~1,000 odorant receptor (OR) genes in the mouse genome, an olfactory sensory neuron (OSN) is thought to express one gene, from one allele. This is reminiscent of immunoglobulin and T-cell receptor genes, which undergo DNA rearrangements in lymphocytes. Here, we test the hypothesis that OR gene choice is controlled by DNA rearrangements in OSNs. Using permanent genetic marking, we show that the choice by an OSN to express an allele of the OR gene *M71* is irreversible. Using *M71*-expressing OSNs as donors for nuclear transfer, we generate blastocysts, embryonic stem (ntES) cell lines and clonal mice. DNA analysis of these cell lines, whose genome is clonally derived from an *M71*-expressing OSN, does not reveal DNA rearrangements or sequence alterations at the *M71* locus. OSNs that differentiate from ntES cells after injection into blastocysts are not restricted to expression of *M71* but can express other OR genes. Thus, *M71* gene choice is irreversible but is reset upon nuclear transfer, and is not accompanied by genomic alterations.

The discovery of OR genes¹ gave rise to the idea that the choice of expression of one allele^{2–4} of one gene⁵ out of a thousand^{6–9} might involve DNA rearrangements, as in immunoglobulin and T-cell receptor genes^{10,11}. A crucial distinction between these types of gene is that the complete coding region of OR genes is already present, preassembled, in the mouse germ line, whereas immunoglobulin and T-cell receptor genes are assembled from gene segments in lymphocyte nuclei by DNA rearrangements. The lack of OSN cell lines expressing an OR gene has precluded a direct search for genomic alterations. Single-cell genomic analysis is technically challenging and offers poor resolution.

Cloning by nuclear transfer has been successful in mouse for a few cell types^{12–16}. Nuclear transfer preserves the DNA rearrangements of immunoglobulin and T-cell receptor loci that occur in the nuclei of B and T lymphocytes¹⁵, and has been proposed as a means of testing the hypothesis of DNA rearrangements of OR genes in OSNs^{7,15}. This method allows the clonal propagation of a nuclear genome as an alternative to establishing a cell line from an OSN. However, as mature OSNs are postmitotic, it is not known whether their nuclei can be reprogrammed to re-enter the cell cycle and instruct the embryonic development of a cloned mouse. Moreover, mature OSNs cannot be obtained as a homogeneous cell type from olfactory epithelium, and OSNs expressing a particular OR gene form an extremely small population of cells within the epithelium. A permanent genetic marker is required to prove that the cloned organisms were derived from a mature OSN. As it is not known whether all mature OSNs express an OR gene, marking should ideally be specific for OSNs that express a predefined OR gene.

Here, we describe three sets of experiments with mature OSNs, permanently marked mature OSNs, and permanently marked OSNs that express the *M71* OR gene. The *M71* promoter has been characterized previously in transgenic mice¹⁷. The permanent marker strategy is based on Cre-mediated activation of a ubiquitously and constitutively expressed green fluorescent protein (GFP), allowing us to identify with certainty the origin of the donor nuclei that supported the establishment of an ntES cell line. From 1,007 successfully reconstructed oocytes, we establish 16 ntES cell lines. We show that OR gene choice is irreversible within an individual

OSN but is reset by nuclear transfer, and does not involve genomic alterations. Similar conclusions are reached in ref. 18.

Mature OSNs as nucleus donors

Mature OSNs can be discerned by specific and high-level expression of the olfactory marker protein (OMP), a protein of unknown function. In OMP–GFP mice (Fig. 1a), in which the OMP-coding region is replaced by that of the GFP, mature OSNs can be identified by their intense green fluorescence¹⁹. In our cell dissociation protocol (see Methods), ~20% of cells from OMP–GFP mice are green fluorescent. We used 411 green-fluorescent cells of OMP–GFP mice successfully as nucleus donors to establish a first set of six embryonic stem (ES) cell lines, which are conventionally referred to as ntES cell lines^{14,20} to reflect the step of nuclear transfer (Table 1). This first set of experiments supports the feasibility of nuclear transfer with mature OSNs, but there is no proof that the nuclei of mature OSNs were the source of these ntES cell lines.

Permanently marked OSNs as nucleus donors

To mark the nuclei of mature OSNs permanently, we generated a novel strain of mice by gene targeting, OMP–Cre, in which the OMP-coding region²¹ is replaced by that of the site-specific Cre recombinase (Fig. 1b). Crossing these mice with Z/EG transgenic reporter mice²² results in Cre-mediated excision of the lacZ-coding sequence, which allows GFP to be expressed from the chicken β -actin promoter. This promoter is ubiquitously and constitutively active in mouse, including in OSNs, such that an OMP–Cre $\times$ Z/EG mouse exhibits a pattern of green fluorescence that is indistinguishable from that of an OMP–GFP mouse. OMP-expressing cells are now permanently marked: their nuclei continue to express GFP protein upon nuclear transfer, regardless of whether the OMP gene is still expressed. This design, developed for lineage tracing, provides retrospective proof that the nucleus came from a cell that was expressing, or had expressed, the OMP gene at the time of observation. From 440 oocytes successfully reconstructed with the nuclei of green-fluorescent cells of OMP–Cre $\times$ Z/EG mice, we cloned green-fluorescent blastocysts (Fig. 1c), and established a second set of seven ntES cell lines (termed OCZ) that are all green fluorescent (Table 1). We generated chimaeric mice by injecting these cells

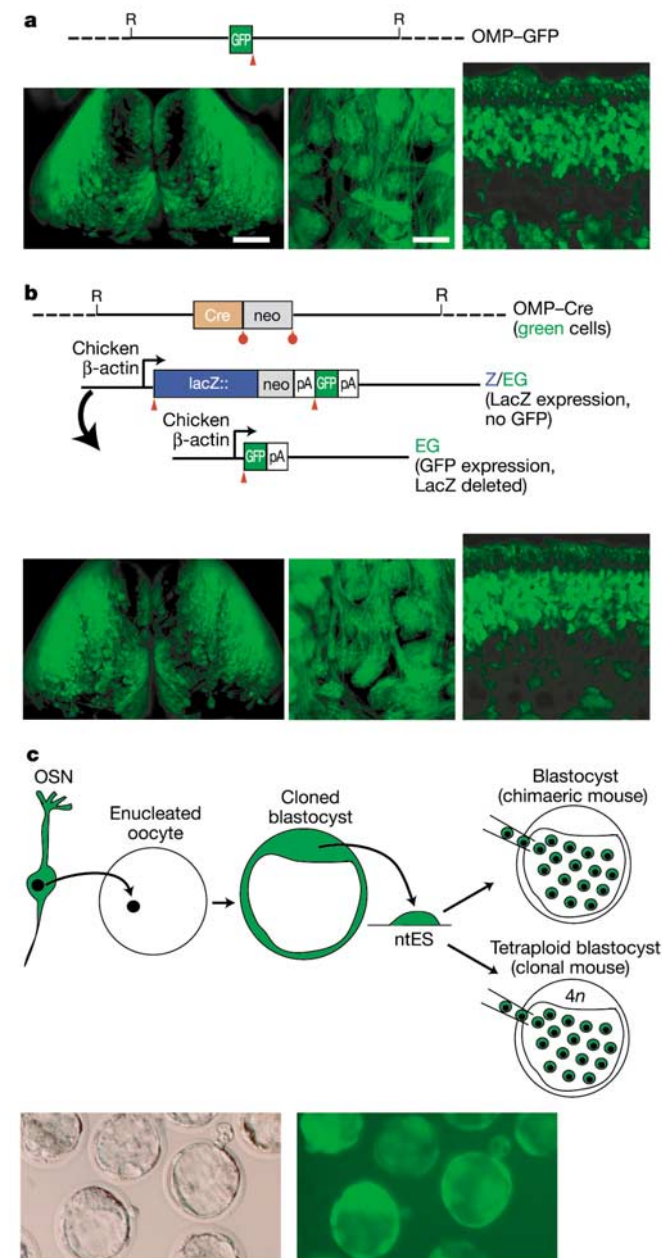


Figure 1 Nuclear transfer with olfactory sensory neurons. **a**, *OMP-GFP* targeted mutation¹⁹. (Red triangles in **a** and **b** are loxP sites, recognition sites for Cre recombinase; R represents EcoRI) Bottom left, wholemount olfactory bulbs. Green-fluorescent glomeruli receive OSN axons. Bottom middle, close-up of bulb showing glomeruli. Bottom right, section through the olfactory epithelium. Mature OSNs are green fluorescent. Scale bars: left, 500 μ m; right, 100 μ m. **b**, Permanent marking. Top, *OMP-Cre* targeted mutation (red dots are FRT sites), the *Z/EG* transgene, and the transgene after Cre-mediated site-specific recombination (GFP, instead of lacZ, is now expressed). Bottom, same views as in **a**, showing that the *OMP-Cre* $\times$ *Z/EG* mouse is phenotypically identical to the *OMP-GFP* mouse. **c**, Experimental design. Green-fluorescent OSNs from a *OMP-Cre* $\times$ *Z/EG* mouse serve as nucleus donors. Reconstructed oocytes develop *in vitro* to blastocysts, from which ntES cell lines are established. Cells are then injected into diploid or tetraploid blastocysts. 'Clonal' mouse denotes the possibility that a small fraction of tetraploid cells contributes to the embryo proper²⁷. Bottom left, Bright-field image of blastocysts after oocyte reconstruction with green-fluorescent OSNs from a *OMP-Cre* $\times$ *Z/EG* mouse. Bottom right, the same blastocysts under fluorescent illumination.

into normal, diploid blastocysts. Two-colour fluorescent *in situ* hybridization (Fig. 2a–c) with probes for *OMP* and *GFP* demonstrates that these ntES cells can give rise to mature, *OMP*-expressing OSNs. Probes for a randomly selected sample of the OR repertoire, *MOR28* (ref. 23) (also known as *MOR244-1*, ref. 6) and *MOR251-4* (ref. 6), co-hybridize in some cells with a probe for *GFP* (Fig. 2d–i). Green-fluorescent OSN axons innervate multiple glomeruli, suggesting that these OSNs express a variety of ORs (Fig. 2j–l).

M71 gene choice is irreversible

Next, we designed a system to generate permanently marked ntES cell lines with nuclei of OSNs that express a predefined and well-characterized OR gene, *M71* (refs 17, 19, 24, 25). We have shown that mini-genes can reproduce the expression features of *M71* in transgenic mice, and have narrowed the regulatory region down to a few hundred nucleotides upstream of the transcription start site (ref. 17, and our unpublished observations). In *M71-IRES-tauGFP* mice²⁵, OSNs expressing the targeted *M71* allele are green fluorescent as a result of co-translational regulation by the internal ribosome entry site (Fig. 3a). We generated a novel mouse strain with a red fluorescent marker based on dsRed²⁶, the *M71-IRES-tauRFP₂* mice (Fig. 3a). In *M71-IRES-tauRFP₂* $\times$ *M71-IRES-tauGFP* mice, which have differentially tagged *M71* alleles, red-fluorescent cells and green-fluorescent cells can be seen in equal numbers in the dorsal epithelium (Fig. 3a). No cell is yellow (from the overlay of red and green fluorescence), illustrating that expression is strictly monoallelic at the time of observation. However, an OSN may switch high-level expression among OR genes during differentiation or throughout its lifetime, while expressing one allele of one OR gene at any moment. If the hypothetical DNA rearrangements would be reversed by this switching (for instance, if induced by a transposon that jumps between promoters, or by gene conversion into an expression cassette), the permanent marker strategy with Cre recombinase is flawed, because GFP expression would not accurately report the OR that is expressed at the time of observation.

To exclude switching for the *M71* gene, we generated a novel strain of mice with an *M71-IRES-Cre* targeted mutation (Fig. 3b), in which Cre recombinase is expressed along with *M71*. By crossing, we obtained mice with two targeted mutations and one transgene: *M71-IRES-tauRFP₂*, *M71-IRES-Cre* and *Z/EG*. Confocal imaging reveals a pattern of red-fluorescent cells and green-fluorescent cells in the epithelium (Fig. 3b) that is identical to the *M71-IRES-tauRFP₂* $\times$ *M71-IRES-tauGFP* cross (Fig. 3a). Likewise, the red-fluorescent and green-fluorescent axons co-converge into the same glomeruli in the olfactory bulb (Fig. 3c–e), indicating that both populations are biologically indistinguishable. Thus, at the time of observation the green-fluorescent cells were expressing the *M71-IRES-Cre* allele instead of having switched expression to another OR gene, and they did not express *M71* transiently from both alleles in their past.

Permanently marked *M71* OSNs as nucleus donors

The irreversibility of *M71* gene choice is essential for interpretation of the third series of nuclear transfer experiments, with green-fluorescent cells of *M71-IRES-Cre* $\times$ *Z/EG* mice. From 156

Donor strains	Reconstructed oocytes	Blastocyst development (%)	Blastocysts used for ntES cell derivation	ntES cell lines (%)
OMP-GFP	411	51 (12.4)	37	6 (16.2)
OMP-Cre $\times$ Z/EG	440	37 (8.4)	25	7 (28.0)
M71-Cre $\times$ Z/EG	156	17 (10.9)	12	3 (25.0)
Total	1,007	105 (10.4)	74	16 (21.6)

successfully reconstructed oocytes, we established three ntES cell lines, termed M71CZ (Fig. 4a–e). Surprisingly, only two cell lines were green fluorescent. The remaining cell line, M71CZ2 (Fig. 4e), was erroneously generated with a nucleus of a cell that was not green fluorescent: genomic DNA of M71CZ2 did not show evidence of Cre-mediated removal of the *loxP*-flanked lacZ-coding region (data not shown). Injection of green-fluorescent M71CZ cells into normal, diploid blastocysts resulted in mice with extensive somatic chimaerism (Fig. 4f, i). We also generated mice by injecting these cells into tetraploid blastocysts. As expected, the cells of these host blastocysts do not contribute to the embryo proper but only to the extra-embryonic tissues, to which, conversely, ES cells cannot contribute (Fig. 4g, h, j). (Because tetraploid cells can make

minor contributions to the embryo proper in certain cases²⁷, the ensuing mice are not necessarily cloned—that is, derived from a single cell—but can be chimaeric; to acknowledge this uncertainty, we refer to these mice as ‘clonal’). Figure 4h shows Harvey, a 3-week-

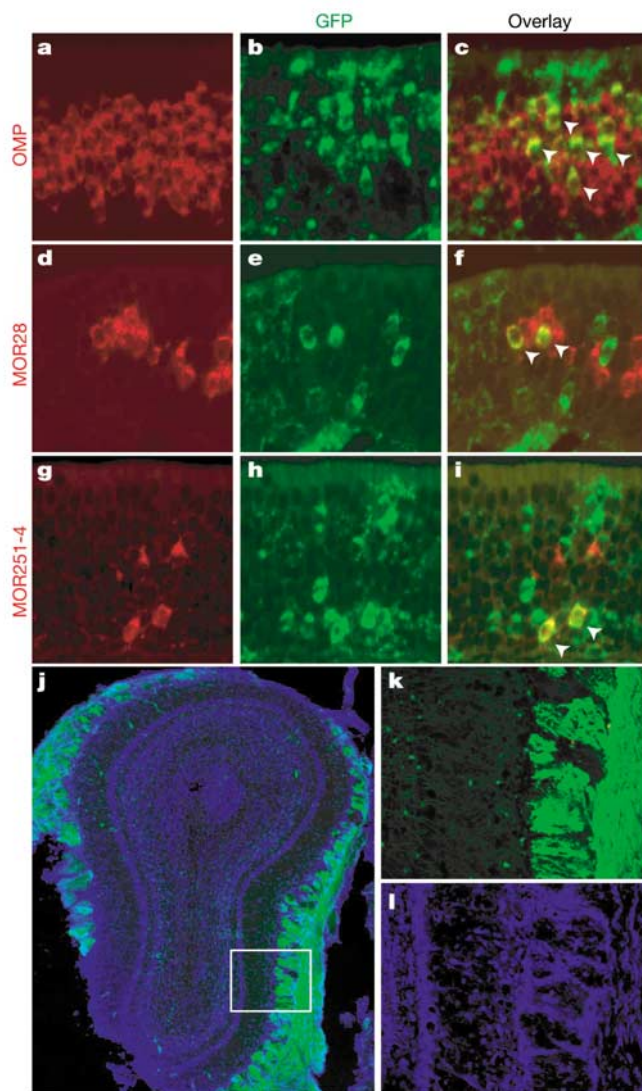


Figure 2 Chimaeras produced with OC21 ntES cells generated from permanently marked, OMP-expressing OSNs. **a–i**, Olfactory epithelium, *in situ* hybridization. Probe for *OMP* in red (**a**); probe for *GFP* in green (**b**); overlay of both (**c**). Arrowheads indicate *OMP*-positive cells derived from the injected ntES cells. Note that some *GFP*-positive cells do not express *OMP*; these are most probably supporting cells. Probe for *MOR28*, expressed in the ventral epithelium, in red (**d**); probe for *GFP* in green (**e**); overlay (**f**). Some *GFP*-positive cells express *MOR28*. Probe for *MOR251-4*, expressed in the intermediate region of the epithelium (**g**); probe for *GFP* (**h**); overlay (**i**). Some *GFP*-positive cells express *MOR251-4*. **j–l**, Olfactory bulb, fluorescence for *GFP* and TOTO-3, a blue nuclear dye. Green-fluorescent axons innervate multiple glomeruli (**j**). Box is magnified in **k** (intrinsic fluorescence of *GFP*) and **l** (TOTO-3, blue). Chimaeras are 2 months old.

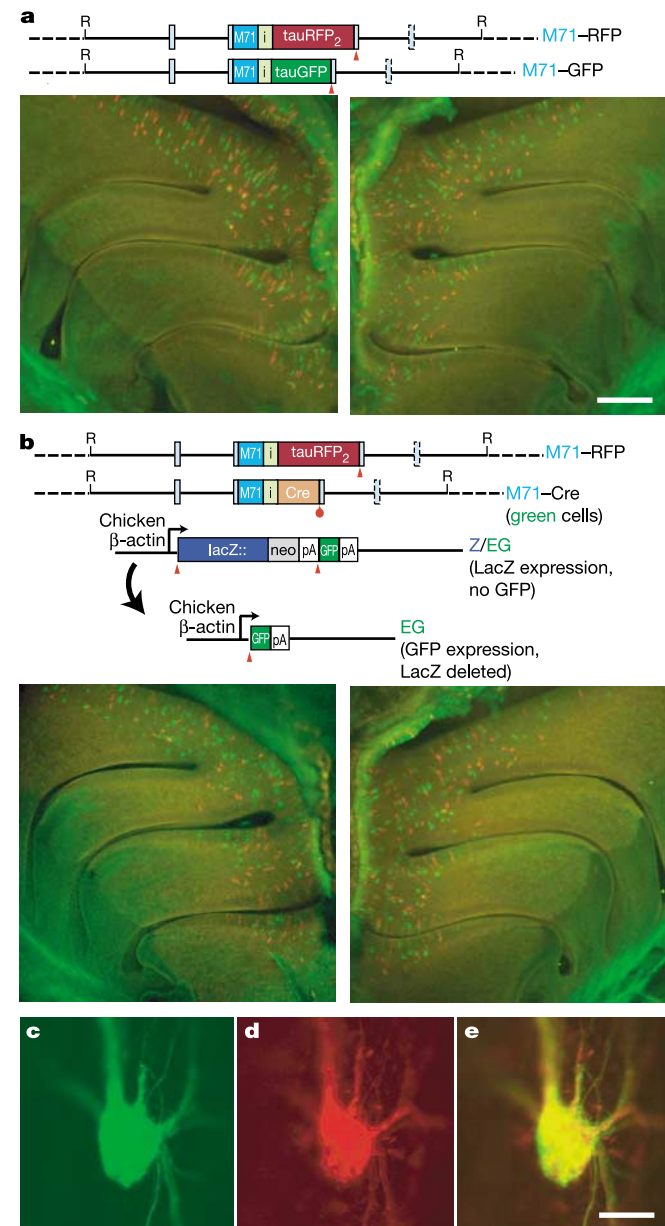


Figure 3 Expression of the *M71* OR gene is strictly monoallelic and does not switch to other OR genes. **a**, Top, *M71*-IRES-*tauRFP2* and *M71*-IRES-*tauGFP* targeted mutations. Bottom, wholemount confocal images of the left and right turbinates of an individual 4-week-old mouse with differentially tagged *M71* alleles. Cells are either red fluorescent ($n = 224$, 48.3%) or green fluorescent ($n = 240$, 51.7%) but not both, and are intermingled in the dorsal epithelium. Scale bar, 500 μ m. **b**, Top, *M71*-IRES-*tauRFP2* and *M71*-IRES-*Cre* targeted mutations, and cross with Z/EG transgenic reporter mouse. Compare with Fig. 1b. Bottom, wholemount confocal images of the left and right turbinates of an individual 4-week-old mouse heterozygous for *M71*-IRES-*tauRFP2*, heterozygous for *M71*-IRES-*Cre*, and hemizygous for Z/EG. Cells are either red fluorescent ($n = 216$, 49.8%) or green fluorescent ($n = 218$, 50.2%) but not both. The pattern is indistinguishable from **a**. **c–e**, Wholemount confocal image of glomerulus of mouse heterozygous for *M71*-IRES-*tauRFP2*, heterozygous for *M71*-IRES-*Cre*, and hemizygous for Z/EG. Green fluorescence (**c**), red fluorescence (**d**), overlay (**e**). The *M71*-IRES-*tauRFP2* and *M71*-IRES-*Cre* alleles are phenotypically identical. Scale bar in **e**, 50 μ m.

old male clonal mouse resulting from injection of M71CZ3 into a tetraploid blastocyst. The generation of chimaeric and clonal mice demonstrates that M71CZ cells do not harbour a genomic alteration that would preclude differentiation into somatic cell types.

The unlimited availability of genomic DNA from the immortal M71CZ ntES cell lines permits a search for DNA rearrangements by conventional methods. We sequenced a region of 1 kilobase in and around the *M71* promoter region that we defined previously in transgenic mice¹⁷, and found no polymorphisms between the *M71-IRES-Cre* (expressing) and wild-type (non-expressing) *M71* allele (Fig. 5a). Additionally, long-range polymerase chain reaction (PCR) between the *M71* promoter region and the *IRES* sequence showed no polymorphisms. Southern blot hybridization with a probe corresponding to the promoter and 5' nontranslated regions did not reveal a band of an altered length indicative of a genomic alteration such as gene conversion (Fig. 5a). We cannot exclude the possibility that a DNA rearrangement took place at distant sites, outside the minimal region for expression that we defined with transgenes¹⁷.

Finally, we performed two-colour *in situ* hybridization in chimaeric and clonal mice produced with green-fluorescent M71CZ ntES cells. Probes for *Cre* (to detect expression of the *M71-IRES-Cre* allele) or *M71* (to detect expression of the *M71-IRES-Cre* and wild-type *M71* alleles, but cross-hybridizing with the *M72* gene) reveal that only a very small subset of GFP-positive cells also express *Cre* (Fig. 5b) or *M71/M72* (Fig. 5c). Moreover, GFP-positive cells express other OR genes, such as *P2* (ref. 21) (Fig. 5d) and *MOR28* (ref. 23) (Fig. 5e). *P2* is expressed in a region of the epithelium that is different from *M71*, as can be assessed by two-colour *in situ* hybridization with *OCAM* (Fig. 5f). Green-fluorescent axons innervate multiple glomeruli, consistent with the expression of any of a variety of ORs in the OSNs (Fig. 5g). Thus, OSNs containing the same nucleus as once belonged to an

M71-expressing cell are not locked in to choose only *M71* for expression but can express other OR genes. More analyses will be required to screen for expression of other OR genes, and to quantify the probability of *M71* gene choice.

Discussion

The efficiency of ntES cell establishment is 1/63 successfully reconstructed oocytes or 1/7 cloned blastocysts in this study, which is comparable to 1/66 oocytes or 1/12 blastocysts in a survey of the mouse ntES literature with other differentiated cell types²⁰.

In OSNs derived from OMP-Cre $\times$ Z/EG ntES cells, the expression of randomly selected genes of the OR repertoire would not be expected if the OSN used for nuclear transfer harboured a genomic alteration in OR gene *X* that instructed ntES-derived OSNs to choose and express invariably OR gene *X*. The argument is that pan-OSN expression of *X* would preclude expression of the rest of the OR repertoire by negative feedback^{28,29}, and would organize the projection of axons to one or a few (gigantic) glomeruli in the olfactory bulb. However, there are several caveats: (1) The donor cells may have expressed OMP in the past but not at the time of observation, and may not correspond to mature OSNs. (2) Some OMP-positive cells may not express any OR gene. Single-cell reverse transcription (RT)-PCR of OSNs yields a cloned OR in 27–83% (average 36%) of cases^{5,30}. The remaining OSNs are either false negative, or do not express any OR gene. (3) The hypothetical genomic alteration in OR gene *X* may be preserved upon nuclear transfer but not be sufficient for expression. As *X* is not known in advance, it can only be identified retrospectively if its expression is greatly increased. (4) Recent results have supported the notion of negative feedback of OR expression^{28,29}, although it is difficult to distinguish it from negative selection against OSNs expressing more than one OR³¹. The mechanism of the presumptive negative feedback may depend on a normal process of OR gene choice. Negative

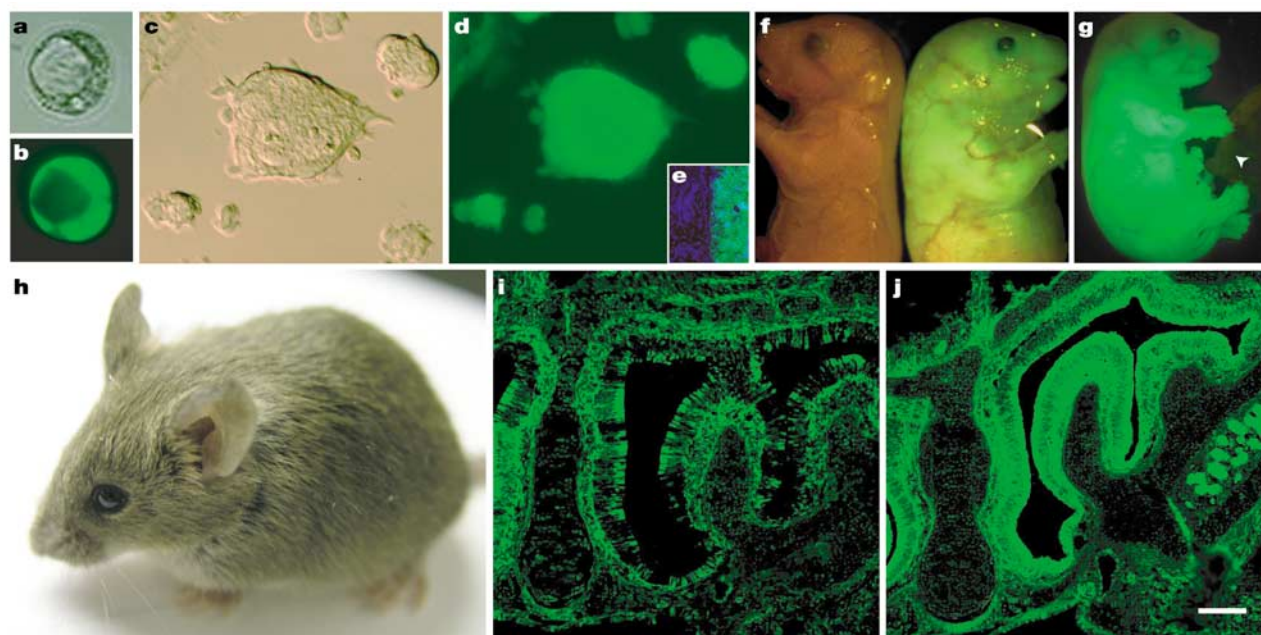


Figure 4 ntES cell lines generated by nuclear transfer from permanently marked, *M71*-expressing OSNs. **a**, Blastocyst, bright-field image. **b**, Same blastocyst, fluorescent image. **c**, M71CZ1 ntES cell colonies in culture, bright-field image. **d**, Same cells, fluorescent image. All cells are strongly green fluorescent. **e**, Comparison of M71CZ2 (not green fluorescent, left) and M71CZ3 (green fluorescent, right). Counterstained with blue nuclear dye TOTO-3. **f**, Normal E19 mouse (left); chimaeric E19 mouse with M71CZ1 ntES cells (right). **g**, Clonal mouse produced by M71CZ1 ntES cell injection into tetraploid

blastocyst. Arrowhead indicates extra-embryonic tissues, which are host-derived and thus not green fluorescent. **h**, Clonal mouse Harvey at 3 weeks. This mouse was generated by injecting M71CZ3 ntES cells (agouti) into a B6D2F1 tetraploid blastocyst (black). **i**, Section through olfactory epithelium of chimaeric mouse produced by diploid blastocyst injection. Not all cells are green fluorescent. **j**, Section through olfactory epithelium of clonal mouse produced by tetraploid blastocyst injection. All cells are green fluorescent. Scale bar, 100 μ m.

feedback may not occur when this choice is bypassed by pan-OSN expression of OR X. (5) OR X may be unable to prevent expression of the limited set of OR genes we tested.

To address these limitations, we carried out experiments with permanently marked OSNs that express a predefined OR, *M71*. Does the marking strategy accurately reflect the OR expressed at the

time of observation? The cross between *M71*–RFP, *M71*–IRES–Cre and *Z/EG* (Fig. 3b, c) demonstrates that once an individual OSN has chosen to express an *M71* allele, expression is irreversible. If switching to another OR gene, transient switching to the other *M71* allele, or biallelic *M71* expression were to occur, green-fluorescent cells would exceed the number of red-fluorescent cells, double-

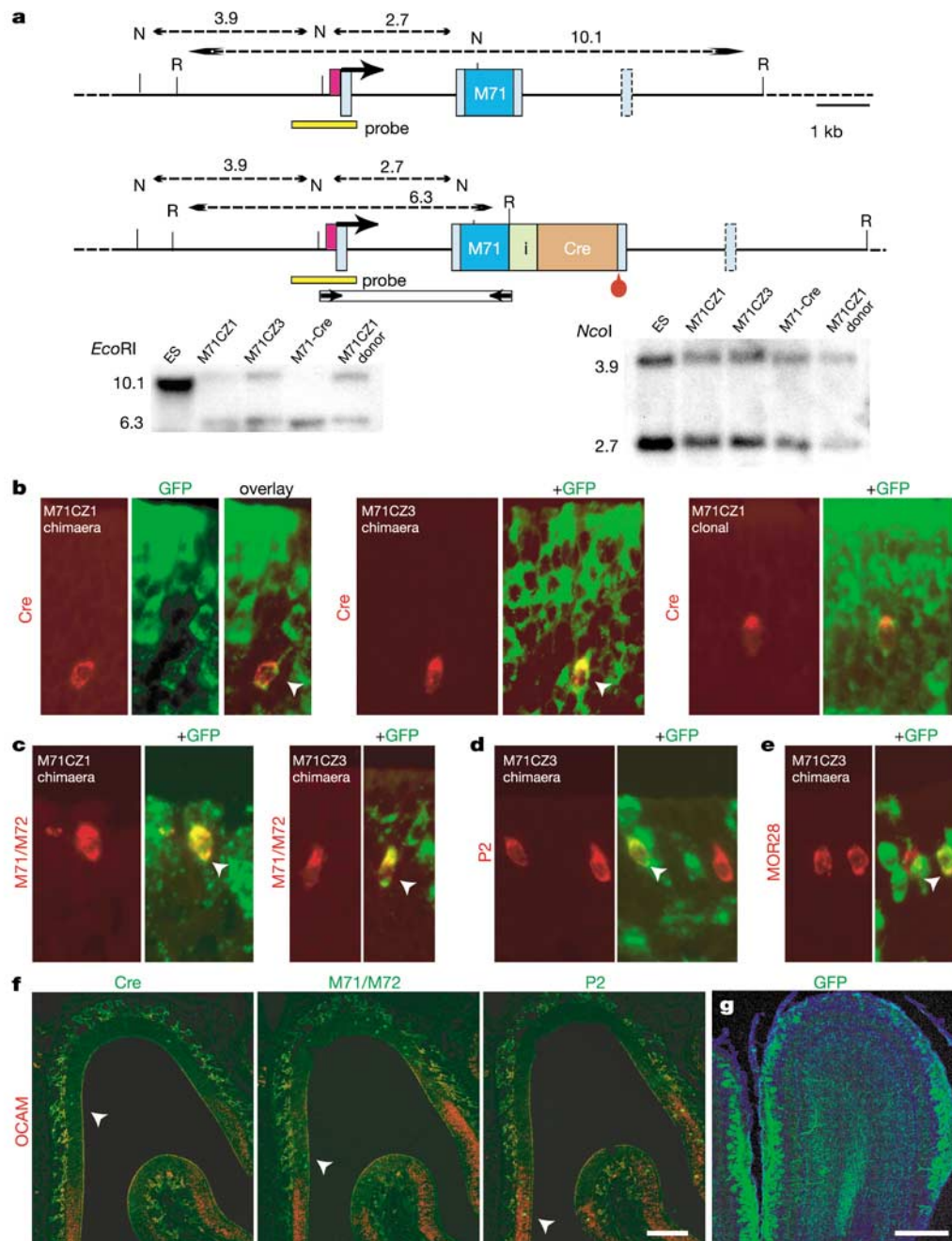


Figure 5 Genomic and phenotypic analysis of *M71CZ* ntES cell lines. **a**, Top, structure of wild-type *M71* locus. Middle, structure of *M71*–IRES–Cre targeted locus. Length of restriction fragments detected by the Southern blot probe (yellow) is indicated. White bar with arrows indicates sequenced region after amplification with long-range PCR. Red box is the minimal upstream region for expression in transgenic mice. R, *EcoRI*; N, *NcoI*. Bottom, Southern blot hybridization. For *EcoRI* digests, the wild-type and *M71*–IRES–Cre allele have bands of different sizes. Lane labelled ‘*M71CZ1* donor’ is DNA from the tail of the mouse that was the nucleus donor for the *M71CZ1* ntES cell line. **b–e**, Two-colour *in situ* hybridization of chimaeric or clonal mice generated by injection of *M71CZ* ntES cells. Probes in red are for *Cre* (detecting the *M71*–IRES–Cre allele) (**b**); *M71/M72* coding

region (**c**); *P2* (**d**); and *MOR28* (**e**). Probe in green in **b** is for *GFP*. Arrowheads in overlays indicate yellow colour, resulting from combined red and green fluorescence. *M71CZ1* chimaera, E19; *M71CZ3* chimaera, 2 weeks; *M71CZ3* clonal, E19. **f**, Two-colour *in situ* hybridization of a 2-week-old chimaera generated by injection of *M71CZ3* ntES cells. Probe in red is for *OCAM*, which is not expressed in the dorsal epithelium. As expected, *Cre* and *M71/M72* probes label cells (arrowheads) in the *OCAM*-negative, dorsal epithelium, whereas *P2* is co-expressed with *OCAM*. Scale bar, 200 μ m. **g**, Section of olfactory bulb of the same mouse as in **f**, counterstained with TOTO-3. Green-fluorescent axons project to multiple glomeruli. Scale bar, 500 μ m.

fluorescent cells would appear, and axons of some green-fluorescent cells (which actually express another OR gene) would project to glomeruli that are distinct from the M71 glomeruli. If a green-fluorescent cell were to cease expression of *M71* and any other OR gene altogether, it would not be expected to survive very long.

These lineage-tracing experiments argue that an immature OSN does not alternate expression between the two *M71* alleles before making its final choice. They further demonstrate that expression of *M71* is irreversible during differentiation and throughout the lifetime of an OSN, at least up to the time of observation. From a practical point of view, they validate the use of the *M71*–IRES–Cre allele as a substitute for the *M71*–IRES–GFP or *M71*–IRES–*tauGFP* allele, with the crucial advantage of a permanent marker.

As outlined in ref. 18, the permanent marker strategy is essential in order to identify with certainty the origin of the donor nucleus that contributed to the generation of a particular cloned animal. Manipulation errors are unavoidable when success is rare and the target cell population is even rarer: in the third and critical set of mice (heterozygous *M71*–IRES–Cre, hemizygous Z/EG), only ~1/10,000 cells in dissociated cell preparations of epithelia is green fluorescent, and 1/52 nuclear transfers generated an ntES cell line. The permanent marker strategy provides a rigorous safeguard against accidental nuclear transfer with any of the remaining 99.99% of cells. Such mistakes would give the same, essentially negative, results (OR gene choice can be reset, and does not involve genomic alterations.) Using *M71*–IRES–GFP or *M71*–IRES–*tauGFP* mice would be inconclusive because such mistakes remain undetected.

Thus, we conclude that expression of the *M71* OR gene is strictly monoallelic and irreversible: expression does not switch to the other *M71* allele or to other OR genes during differentiation and throughout the lifetime of an OSN. *M71* gene choice is not instructed by DNA rearrangements, assuming that these genomic alterations would have been preserved upon nuclear transfer and ntES cell derivation. Thus, OR gene expression is unlike that of immunoglobulin and T-cell receptor genes. □

Methods

Gene targeting

ES cells of cell line E14 (of 129 origin) were used³². *OMP* and *M71* targeting vectors have been described^{21,25}. The *Cre* coding sequence was excised from *pBS185* (ref. 33). *PacI* cassettes inserted were *Cre* for *OMP* and *IRES–Cre* along with a *neo*-selectable marker flanked by *FRT* sites. *FNF*. Flp-mediated excision of *M71–Cre* was carried out *in vivo* by crossing to *Actin–flp* transgenic mice and subsequent outcrossing of the *flp* transgene³⁴. The ES clone name for *OMP–Cre* is F60 and for *M71–Cre* clone G12. *M71–RFP* (clone TD71-8) is the *M71–IRES–tauRFP* mutation with the *neo*-selectable marker removed using *ACN*/by self-excision in the germ cells of male chimaeras³⁵. RFP (dsRed-1)²⁶ is from Clontech; RFP2 is a dsRed dimer linked by seven amino acids, VDPVAT. Mice are in a mixed 129 × C57BL/6 background. Z/EG mice²² were purchased from the Jackson Laboratory.

Dissociation of OSNs

Olfactory epithelia were dissected from 2-week-old (*OMP–Cre*), and 2–4-week-old mice (*M71–Cre*), hemizygous for Z-E/G and heterozygous for the targeted mutation. Tissue was dissected in Ringer's solution, and dissociated with forceps and razor blades; no enzymes were used. Cells were washed three times with HEPES–CZB medium and suspended in HEPES–CZB medium containing 3% (w/v) polyvinylpyrrolidone³⁶.

Nuclear transfer and embryo culture

This was performed as described¹². Briefly, metaphase II-arrested oocytes were collected from superovulated B6D2F1 females (8–10 weeks) and cumulus cells were removed using hyaluronidase. Oocytes were cultured in CZB medium supplemented with 5.5 mM glucose at 37 °C under 5% CO₂ in air. Enucleation was performed in a droplet of HEPES–CZB medium containing 5 µg ml^{−1} cytochalasin B using a blunt Piezo-driven pipette. After enucleation, the spindle-free oocytes were washed extensively and kept in CZB medium up to 2 h before nucleus injection. Green-fluorescent donor cells were picked up and aspirated in and out of the injection pipette to remove the cytoplasmic material. Each nucleus was injected into an enucleated oocyte using Piezo microinjection. The reconstructed oocytes were cultured in CZB medium for 1–3 h and then were activated for 5–6 h in calcium-free CZB medium containing 10 mM Sr²⁺ and 5 µg ml^{−1} cytochalasin B. Following activation, all reconstructed embryos were cultured in KSOM medium with amino acids at 37 °C under 5% CO₂ in air.

ES cell derivation

The reconstructed embryos were cultured in KSOM medium with amino acids for 3.5–4 days to reach the blastocyst stage. Expanded or hatched blastocysts were selected to establish ntES cell lines as described¹⁴. Each blastocyst was transferred into one well of a 96-well plate seeded with mouse-strain ICR embryonic fibroblast feeders in ES medium supplemented with 20% knockout serum replacement, 1,500 U ml^{−1} leukaemia inhibitory factor (LIF) and 50 µM of the MEK1 inhibitor (PD98059) (ref. 37) after the zona pellucida was removed using acidic Tyrode's solution. After 4–7 days in culture, colonies were trypsinized and transferred to a 96-well plate with a fresh feeder layer in fresh medium. Clonal expansion of ntES cells proceeded from 48-well plates to 6-well plates with feeder cells and then to gelatinized 25-cm² flasks for routine culture in ES medium, with 15% fetal calf serum (FCS) and 1,000 U ml^{−1} LIF.

Injection of ntES cells into blastocysts

Diploid blastocysts were collected from the uterus of E3.5 superovulated C57BL/6 females and kept in KSOM medium with amino acids until ntES cell injection. To obtain tetraploid blastocysts, two-cell embryos from E2 superovulated B6D2F1 females were electrofused to produce one-cell tetraploid embryos. Two-cell embryos were aligned by alternating current (a.c.) in 0.3 M mannitol solution, and a single direct current (d.c.) pulse of 1,000 V cm^{−1} for 20 µs was applied. After electrofusion, embryos were returned to KSOM with amino acids and fused embryos were cultured for two days to reach the blastocyst stage. For blastocyst injection ES cells were trypsinized, resuspended in DMEM without LIF, and kept on ice. A flat tip microinjection pipette was used for ES cell injection. One hundred ES cells were picked up in the end of the injection pipette and injected into ten blastocysts. The blastocysts were kept in KSOM with amino acids until embryo transfer. Ten injected blastocysts were transferred into each uterine horn of 2.5-days-postcoitum pseudopregnant ICR females. Pregnant recipients with tetraploid embryos were subjected to caesarean section on day 19.5 of gestation.

Sequence analysis

We have shown that 161 base pairs (bp) upstream of the *M71* transcription start site is sufficient for the recapitulation of *M71* expression in transgenic mice (ref. 17, and our unpublished observations). Oligonucleotides were designed to amplify ±500 bp of the *M71* transcription start site. Long-range PCR was performed from −241 bp until the *IRES* sequence, which immediately follows the *M71* coding region, with oligonucleotides 5'–CGGCTC TGTGTGATGCTAAATGTTTCATTGCC and 3'–GCGGAATCTTAATTAAACATCAAAGACTTTTC. PCR products were directly sequenced.

In situ hybridization

Antisense digoxigenin- and/or fluorescein-labelled complementary RNA probes prepared from cloned segments of complementary DNAs encoding *GFP* (full coding sequence), *OMP* (nucleotides (nt) 820–2891 from U01213), *Cre* (full coding sequence), *M71* (nt 64–930 from AF281061), *P2* (nt 54–948 from AF247657), *MOR251-4* (nt 483–895 from NG002083), *MOR28* (ref. 23), *OCAM* (nt 496–2002 from AF001287). Sections (10 µm) were cut from the olfactory epithelium. Two-colour fluorescent *in situ* hybridization was performed as described^{38,39}. The digoxigenin-labelled probe was visualized with HNPP/FAST RED, and the fluorescein-labelled probe was visualized with Alexa 488. Sections were analysed using a confocal microscope Zeiss LSM 510.

Received 22 January; accepted 17 February 2004; doi:10.1038/nature02433.

- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
- Strotmann, J. et al. Local permutations in the glomerular array of the mouse olfactory bulb. *J. Neurosci.* **20**, 6927–6938 (2001).
- Ishii, T. et al. Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurons. *Genes Cells* **6**, 71–78 (2001).
- Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
- Zhang, X., Rodriguez, I., Mombaerts, P. & Firestein, S. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics* published online 4 December 2003 (doi:10.1016/j.ygeno.2003.10.009).
- Mombaerts, P. Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**, 707–711 (1999).
- Mombaerts, P. How smell develops. *Nature Neurosci.* **4**, 1192–1198 (2001).
- Mombaerts, P. Genes and ligands for odorant, vomeronasal and taste receptors. *Nature Rev. Neurosci.* **5**, 263–278 (2004).
- Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl Acad. Sci. USA* **73**, 3628–3632 (1976).
- Tonegawa, S. Somatic generation of immune diversity *in vitro*. *Cell. Dev. Biol.* **24**, 235–265 (1988).
- Wakayama, T., Perry, A. C. F., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
- Wakayama, T., Rodriguez, I., Perry, A. C., Yanagimachi, R. & Mombaerts, P. Mice cloned from embryonic stem cells. *Proc. Natl Acad. Sci. USA* **96**, 14984–14989 (1999).
- Wakayama, T. et al. Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. *Science* **292**, 740–743 (2001).
- Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035–1038 (2002).
- Hochedlinger, K. & Jaenisch, R. Nuclear transplantation, embryonic stem cells, and the potential for cell therapy. *N. Engl. J. Med.* **349**, 275–286 (2003).
- Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M. & Mombaerts, P. Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron* **35**, 681–696 (2002).

18. Eggan, K. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428**, 44–49 (2004).
19. Potter, S. M. *et al.* Structure and emergence of specific olfactory glomeruli in the mouse. *J. Neurosci.* **21**, 9713–9723 (2001).
20. Mombaerts, P. Therapeutic cloning in the mouse. *Proc. Natl Acad. Sci. USA* **100**, 11924–11925 (2003).
21. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
22. Novak, A., Guo, C., Yang, W., Nagy, A. & Lobe, C. G. Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. *Genesis* **28**, 147–155 (2000).
23. Tsuboi, A. *et al.* Olfactory neurons expressing closely linked and homologous odorant receptor genes tend to project their axons to neighboring glomeruli on the olfactory bulb. *J. Neurosci.* **19**, 8409–8418 (1999).
24. Zheng, C., Feinstein, P., Bozza, T., Rodriguez, I. & Mombaerts, P. Peripheral olfactory projections are differentially affected in mice deficient in a cyclic nucleotide-gated channel subunit. *Neuron* **26**, 81–91 (2000).
25. Bozza, T., Feinstein, P., Zheng, C. & Mombaerts, P. Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* **22**, 3033–3043 (2002).
26. Matz, M. V. *et al.* Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnol.* **17**, 969–973 (1999).
27. Nagy, A. *et al.* Embryonic stem cells alone are able to support fetal development in the mouse. *Development* **110**, 815–821 (1990).
28. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor–one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
29. Lewcock, J. W. & Reed, R. R. A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl Acad. Sci. USA* **101**, 1069–1074 (2004).
30. Hamana, H., Hirono, J., Kizumi, M. & Sato, T. Sensitivity-dependent hierarchical receptor codes for odors. *Chem. Senses* **28**, 87–104 (2003).
31. Mombaerts, P. Odorant receptor gene choice in olfactory sensory neurons: the one receptor–one neuron hypothesis revisited. *Curr. Opin. Neurobiol.* **14**, 31–36 (2004).
32. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 (1987).
33. Sauer, B. & Henderson, N. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol.* **2**, 441–449 (1990).
34. Dymecki, S. M. Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 6191–6196 (1996).
35. Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germ line. *Genes Dev.* **13**, 1524–1528 (1999).
36. Chung, Y. G., Mann, M. R., Bartolomei, M. S. & Latham, K. E. Nuclear-cytoplasmic “tug of war” during cloning: effects of somatic cell nuclei on culture medium preferences of preimplantation cloned mouse embryos. *Biol. Reprod.* **66**, 1178–1184 (2002).
37. Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signaling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* **210**, 30–43 (1999).
38. Hirota, S. *et al.* Localization of mRNA for c-kit receptor and its ligand in the brain of adult rats: an analysis using *in situ* hybridization. *Mol. Brain Res.* **15**, 47–54 (1992).
39. Ishii, T., Hirota, J. & Mombaerts, P. Combinatorial coexpression of neural and immune multigene families in mouse vomeronasal sensory neurons. *Curr. Biol.* **13**, 394–400 (2003).

Acknowledgements We thank K. Amuluru, J. Rosenberg and J. Wu for technical assistance in preparing dissociated cells, T. Bozza for helping to construct M71-IRES-tauRFP₂ mice, D. Wen for collaborating to set up the tetraploid blastocyst injection technique, H. Sakano for the MOR28 probe, and R. Gao and K. Latham for advice about nuclear transfer. T.I. was supported by a postdoctoral fellowship from the Human Frontier Science Program organization. P.M. is grateful for grant support from the NIH/NIDCD.

Competing interests statement The authors decline to provide information about competing financial interests.

Correspondence and requests for materials should be addressed to P.M. (peter@rockefeller.edu).

Retrograde spins of near-Earth asteroids from the Yarkovsky effect

A. La Spina¹, P. Paolicchi², A. Kryszczyńska³ & P. Pravec⁴

¹Dipartimento di Matematica, ²Dipartimento di Fisica, Università di Pisa, via Buonarroti 2, I-56127 Pisa, Italy

³Astronomical Observatory, Adam Mickiewicz University, Słoneczna 36, 60286 Poznań, Poland

⁴Astronomical Institute AS CR, Fricova 1, CZ-25165 Ondřejov, Czech Republic

Dynamical resonances in the asteroid belt are the gateway for the production of near-Earth asteroids¹ (NEAs). To generate the observed number of NEAs, however, requires the injection of many asteroids into those resonant regions. Collisional processes have long been claimed as a possible source^{1–3}, but difficulties with that idea have led to the suggestion that orbital drift arising from the Yarkovsky effect^{4–7} dominates the injection process^{8–10}. (The Yarkovsky effect is a force arising from differential heating—the ‘afternoon’ side of an asteroid is warmer than the ‘morning’ side.) The two models predict different rotational properties of NEAs: the usual collisional theories² are consistent with a nearly isotropic distribution of rotation vectors, whereas the ‘Yarkovsky model’ predicts an excess of retrograde rotations. Here we report that the spin vectors of NEAs show a strong and statistically significant excess of retrograde rotations, quantitatively consistent with the theoretical expectations of the Yarkovsky model.

The spin vector of an asteroid is determined from the rotation period (available for ~1,000 asteroids¹¹) and two orientation angles. This latter information is usually given in terms of ecliptic latitudes and longitudes of the rotation axes, and is reliable for ~100 objects^{12,13}. The results for both NEAs and main belt asteroids are displayed in Fig. 1.

The distribution of spin vectors (poles) among main belt asteroids shows a moderate excess of prograde spins (confirming previous findings^{11,14}, and perhaps arising from the original rotational properties), and a strong absence of vectors pointing close to the ecliptic plane. A χ^2 analysis shows that the probability of a statistical fluctuation resulting in such a distribution is 0.07%. The NEAs exhibit a strong excess of retrograde rotation (14 objects) compared to prograde (7) bodies. Almost 50% of NEAs are in the first bin (of seven), so the ecliptic plane can be considered depopulated. A statistical analysis confirms that the likelihood of the NEA and main belt distributions being similar has a probability of about 1%. A similar value (~3%) holds for the probability of the NEA distribution being consistent with a flat distribution.

The depopulation of poles close to the ecliptic has been puzzling. When it was first noted, with only about one-quarter of the present sample and with the observational techniques of the 1980s, a selection effect was claimed as a possible and partial explanation¹⁴. This suggestion, however, cannot work with the present sample. One might connect this anomaly with the pole clustering—far from the ecliptic—arising from the YORP effect^{15,16} (a side consequence of the Yarkovsky effect), but YORP is effective only for bodies smaller than ~40 km; in our main belt sample most of the bodies are larger (often far larger). This point deserves further analysis.

A dynamical selection effect could explain the excess of retrograde spinning NEAs. There is no obvious reason why NEAs have to become retrograde, but there is a reason why retrograde objects are favoured to become NEAs, if the injection of asteroids into NEA-like orbits arises from the Yarkovsky effect.

The effect changes the orbital semimajor axis of an asteroid, which consequently migrates towards one of the major resonances (3:1 and ν_6 (ref. 9) have been considered to be the preferential channels to enter NEA-formation trackways). Among the NEAs in

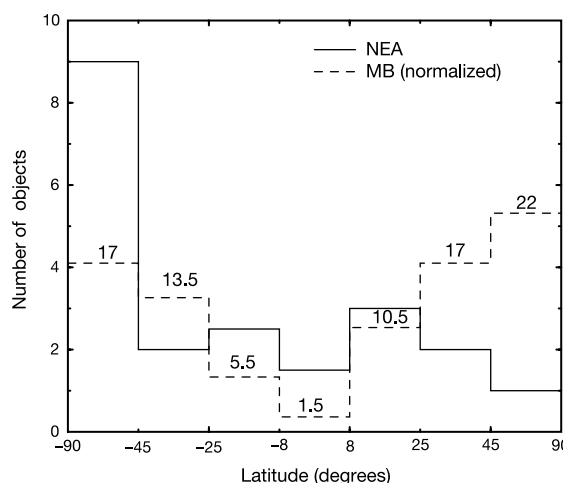


Figure 1 Latitude distributions for main belt asteroids and NEAs. The histogram shows (solid line) the distribution of the ecliptic latitude for the poles of 21 NEAs (the database¹² has been updated with a few new data¹³). The zones referred to in the abscissa represent^{11,14} approximately equal surfaces on the λ, β sphere; region 1 means retrograde, from -90° to -45° , 2 means -45° to -25° , 3 means -25° to -8° , 4 means -8° to 8° (prograde) and so on. The distribution is compared to that computed for 87 main belt asteroids (dashed line, MB), which has been normalized to the same number of objects. The numbers above the dashed segments correspond to the real number of main belt asteroids in the bin. The contradictory determinations present in the database¹² have been omitted; the ambiguous ones (rare among NEAs, but very common among main belt objects, owing to a partial sampling of phase and aspect angles) have been weighted as 1/2. Among the NEAs in the first bin only two bodies are taxonomically inconsistent with type S, thus supporting our suggestion that most of them come out from the inner part of the belt. However, several (7/12) S-type bodies (or similar) are also present in the other bins. Note also that the taxonomically S-type NEAs differ in subtle ways from the corresponding main belt ones¹⁷. In principle, the latitudes might be taken with respect to the orbital plane of every individual asteroid: the results should be very similar (15 retrograde bodies among NEAs instead of 14). However, the present NEA orbits have been severely affected by the resonant interactions and—eventually—by close encounters. They are not similar or trivially correlated to the original ones, which were affected by the Yarkovsky drift.

our list 19 objects are smaller than 10 km, so the Yarkovsky drift may have been very effective. Note also that the drift is more effective when the spin axis is not too close to the orbital plane.

The ν_6 resonance is located at the inner edge of the main belt and therefore can be reached only by objects evolving inward, towards the Sun. The only objects that can enter the ν_6 resonance through the Yarkovsky effect are those with retrograde spins. All the other important main belt escape channels, such as the 3:1 resonance, can be reached by objects drifting either inward and outward (that is, prograde and retrograde spinning).

It is possible to estimate the fraction of NEAs coming from various source regions in the main belt^{8–10}. Some computations^{8,9} suggest that a percentage $P_{\nu_6} \approx 37\%$ of all NEAs (with asteroidal absolute magnitude $H < 18$) came from the ν_6 resonance, with the rest coming from a variety of different source regions (among them the 3:1 resonance). The fraction coming from ν_6 decreases to about 30% according to a different model¹⁰. Thus the theoretical ratio between NEAs with retrograde/prograde spin can be estimated to be:

$$N_r/N_p \approx (1 + P_{\nu_6})/(1 - P_{\nu_6}) \approx 2 \pm 0.2$$

The present data give $2^{+1}_{-0.7}$, with a surprisingly good fit.

We conclude that the distribution of spin vectors strongly supports the key role of Yarkovsky-induced drift in the injection of main belt asteroids into the resonant regions. In principle, the rotation could be used with other observational data to determine the origin of every individual near-Earth object. □

Received 5 December 2003; accepted 9 February 2004; doi:10.1038/nature02411.

- Greenberg, R. & Nolan, M. C. in *Asteroids II* (eds Binzel, R. P. et al.) 778–804 (Arizona Univ. Press, Tucson, 1989).
- Davis, D. R., Weidenschilling, S. J., Farinella, P., Paolicchi, P. & Binzel, R. P. in *Asteroids II* (eds Binzel, R. P. et al.) 805–826 (Arizona Univ. Press, Tucson, 1989).
- Zappalà, V., Cellino, A., Di Martino, M., Migliorini, F. & Paolicchi, P. Maria's family: physical structure and possible implications for the origin of giant NEAs. *Icarus* **129**, 1–20 (1997).
- Vokrouhlický, D. Diurnal Yarkovsky effect as a source of mobility of meter-sized asteroidal fragments. I. Linear theory. *Astron. Astrophys.* **335**, 1093–1100 (1998).
- Vokrouhlický, D. A complete linear model for the Yarkovsky thermal force on spherical asteroid fragments. *Astron. Astrophys.* **344**, 362–366 (1999).
- Vokrouhlický, D. & Farinella, P. The Yarkovsky seasonal effect on asteroidal fragments: A nonlinearized theory for the plane-parallel case. *Astron. J.* **116**, 2032–2041 (1998).
- Spitale, J. & Greenberg, R. Numerical evaluation of the general Yarkovsky effect: Effects on semimajor axis. *Icarus* **149**, 222–234 (2001).
- Bottke, W. F., Jedicke, R., Morbidelli, A., Petit, J. & Gladman, B. Understanding the distribution of near-Earth asteroids. *Science* **288**, 2190–2194 (2000).
- Bottke, W. F. et al. Debaised orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* **156**, 399–433 (2002).
- Morbidelli, A. & Vokrouhlický, D. The Yarkovsky-driven origin of near-Earth asteroids. *Icarus* **163**, 120–134 (2003).
- Pravec, P., Harris, A. W. & Michalowski, T. in *Asteroids III* (eds Bottke, W. F. Jr et al.) 113–122 (Arizona Univ. Press, Tucson, 2002).
- Kryszczyńska, A. *Database of Asteroid Spin Vectors Determinations*. (<http://www.astro.amu.edu.pl/Science/Asteroids/>) accessed 23 October 2003.
- Kaasalainen, M. et al. Photometry and models of eight near-Earth asteroids. *Icarus* **167**, 178–196 (2004).
- Magnusson, P. et al. in *Asteroids II* (eds Binzel, R. P. et al.) 66–97 (Arizona Univ. Press, Tucson, 1989).
- Rubincam, D. P. Radiative spin-up and spin-down of small asteroids. *Icarus* **148**, 2–11 (2000).
- Vokrouhlický, D., Nesvorný, D. & Bottke, W. F. The vector alignments of asteroid spins by thermal torques. *Nature* **425**, 147–151 (2003).
- Cellino, A., Bus, S. J., Doressoundiram, A. & Lazzaro, D. in *Asteroids III* (eds Bottke, W. F. Jr et al.) 633–643 (Arizona Univ. Press, Tucson, 2002).

Acknowledgements This work was supported by MIUR (Italy), KBN (Poland) and the Academy of Sciences (Czech Republic). We thank D. Vokrouhlický and S. Shore for comments and suggestions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to P. Paolicchi (paolicchi@df.unipi.it).

Strain-induced metal–insulator phase coexistence in perovskite manganites

K. H. Ahn, T. Lookman & A. R. Bishop

Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

The coexistence of distinct metallic and insulating electronic phases within the same sample of a perovskite manganite^{1–6}, such as $\text{La}_{1-x}\text{Pr}_x\text{Ca}_x\text{MnO}_3$, presents researchers with a tool for tuning the electronic properties in materials. In particular, colossal magnetoresistance⁷ in these materials—the dramatic reduction of resistivity in a magnetic field—is closely related to the observed texture owing to nanometre- and micrometre-scale inhomogeneities^{1–6,8}. Despite accumulated data from various high-resolution probes, a theoretical understanding for the existence of such inhomogeneities has been lacking. Mechanisms invoked so far, usually based on electronic mechanisms and chemical disorder^{9–11}, have been inadequate to describe the multiscale, multiphase coexistence within a unified picture. Moreover, lattice distortions and long-range strains^{12,13} are known to be important in the manganites¹⁴. Here we show that the texturing can be due to the intrinsic complexity of a system with strong coupling between the electronic and elastic degrees of freedom. This leads

to local energetically favourable configurations and provides a natural mechanism for the self-organized inhomogeneities over both nanometre and micrometre scales. The framework provides a physical understanding of various experimental results and a basis for engineering nanoscale patterns of metallic and insulating phases.

Existing theory^{9–11} relies predominantly on two very different mechanisms to explain inhomogeneities in manganites: electronic phase separation for nanometre-scale inhomogeneity, and the effects of disorder for micrometre-scale inhomogeneity. Thus two separate mechanisms are invoked to explain two different length scales, in contrast to some experiments^{1,2,6,15}. Nanometre-scale inhomogeneities are typically predicted from simulation results of models^{9,10} in which the charge density is an order parameter. However, these do not include the long-range Coulomb interaction, and even if it were included, it is not clear whether the inhomogeneity would disappear or whether the system would form polarons or nanometre-scale droplets. Similarly, the random-field Ising models^{9,10} proposed for micrometre-scale domains do not capture long-range elastic interactions that are known to be crucial for manganites. Where phenomenological long-range fields are included¹¹, it is assumed that the metal and insulator states have exactly the same energy in the average random field, even though the two states are not related by symmetry. This leads to a homogeneous phase, contrary to the desired goal of obtaining inhomogeneous states. Although it has been recognized^{16–18} that the coupling of spin, charge and lattice distortions together with aspects of nonlinearity are necessary for understanding inhomogeneity in perovskite manganites, few calculations exist that demonstrate how coexistence between metallic and insulating phases arises.

To capture the salient aspects of strong electron–lattice coupling¹⁴ in manganites, we propose a model in which the phase with short- and long-wavelength lattice distortions is insulating, and that without lattice distortion is metallic. Our generic model does not explicitly include other details of perovskite manganites, such as the origin of the short-wavelength lattice distortions (which would be particularly related to oxygen displacements) or magnetic properties¹⁴. However, we consider that it is the structural aspect that primarily causes the multiphase coexistence, an aspect common to purely strain-based materials such as martensites^{19,20} which undergo solid–solid phase transformations. The structural templates then drive novel electronic, magnetic and optical properties in manganites. The explicit inclusion of lattice degrees of freedom, together with the effects of electron–lattice coupling and the use of a single mechanism to obtain both nanometre- and micrometre-scale inhomogeneities, distinguishes our model from other existing models^{9–11}. The multiphase coexistence originates from lattice degrees of freedom rather than charge density, so including the Coulomb interaction would not change the qualitative features of our results, particularly if the charge fluctuations on Mn sites are small or screened, as suggested recently²¹. Thus our microscopic model describes submicrometre texture that does not necessarily rely on substantial fluctuations in charge density even at the atomic scale. Separation of lattice distortions into short (s_x and s_y in Fig. 1a) and long (e_1 , e_2 and e_3 in Fig. 1a) wavelength modes and constraints amongst them, which mediate long-range anisotropic elastic interactions, have been developed and tested on simple examples^{12,13}, and are used as a technical tool in our current model for manganites. Our model also provides a basis for understanding other features observed in manganites, such as precursor short-range ordering⁸ and quasi-elastic scattering^{22,23} near the phase-transition temperature, hysteretic and glassy dynamics^{22,24}, metastability²⁵ and a photo-induced insulator–metal transition^{26,27}.

The lattice distortions in manganites closely follow the state of the outermost shell (e_g) electrons on Mn ions through a Jahn–Teller coupling¹⁴. If an e_g electron is localized at a Mn site in the insulating phase, the symmetry of the surrounding oxygen octahedron is

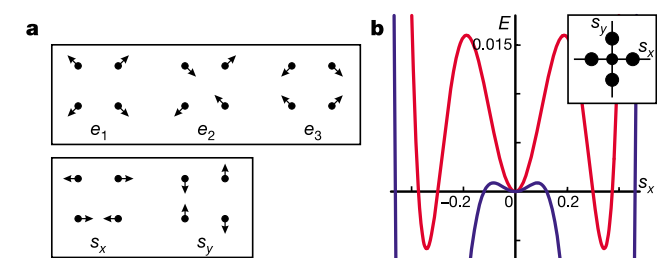


Figure 1 Modes and energy landscape. **a**, Atomic-scale lattice distortion modes (positive) for a monatomic square lattice in two-dimensional (2D) space¹². **b**, Energy landscape along $s_y = 0$ in $s_x - s_y$ plane for two different positive values of the harmonic modulus for short-wavelength distortion, B , in the energy expression below. The global minimum of the blue curve is at $E = -0.11$. The solid circles in the inset schematically represent the locations of local minima in the $s_x - s_y$ plane. We consider the situation in which the undistorted state itself is a local energy minimum and the distorted state is the global energy minimum. For $B < 0$, the resultant double-well energy potential, for which the undistorted phase is unstable, would merely yield as inhomogeneity a line (a plane in three dimensions, 3D) of atoms instead of submicrometre-sized domains observed by electron microscopy². The energy expression for our 2D model for a square lattice is $E = E_{\text{long}} + E_{\text{short}} + E_{\text{coupling}}$, where $E_{\text{long}} = \sum_{\text{sites}} A_1 e_1^2/2 + A_2 e_2^2/2 + A_3 e_3^2/2$, and $E_{\text{short}} = \sum_{\text{sites}} B(s_x^2 + s_y^2)/2 + G_1(s_x^4 + s_y^4)/4 + G_2 s_x^2 s_y^2/2 + H_1(s_x^6 + s_y^6)/6 + H_2 s_x^2 s_y^2(s_x^2 + s_y^2)/6$, and $E_{\text{coupling}} = \sum_{\text{sites}} C_3(s_x^2 - s_y^2)e_3$. E_{long} contains the harmonic energy for dilatation (e_1), shear (e_2) and deviatoric (e_3) long-wavelength modes, which lead to an anisotropic long-range interaction²⁹. The symmetry-allowed energy terms for short-wavelength modes (s_x, s_y) are considered up to sixth order to include features associated with first-order phase transitions. E_{coupling} represents the coupling between short- and long-wavelength modes, where the positive C_3 is the strength of this coupling. In the homogeneous phase, all modes are independent and E can be minimized separately for each mode¹². Such separate minimization gives $e_1 = e_2 = 0$ and $e_3 = -C_3(s_x^2 - s_y^2)/A_3$, which renormalize the fourth-order coefficients and provide the condition to have both distorted and undistorted phases as local minimum states in the $s_x - s_y$ plane. Our model can be extended to study the effects of other perturbations, such as the size distribution of Re/Ak ions^{15,30}, substrate-induced strains in thin films, external stress or grain boundaries in polycrystals^{1,16,19}.

lowered from cubic to tetragonal. At low temperatures, the distorted octahedra stack in particular patterns, often referred to as charge and orbital ordering¹. For example, in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$, the long Mn–O bonds of the elongated octahedra form a zigzag pattern in the x – y plane²⁸, giving rise to short-wavelength lattice distortions. The stacking of the short Mn–O bonds along the z -direction is responsible for the uniform (or long-wavelength) tetragonal (more accurately orthorhombic) distortion. Such lattice distortions are absent in the metallic phase because the e_g electrons are delocalized.

The coupled short- and long-wavelength modes in the cubic anharmonic elastic energy give rise naturally to an energy ‘landscape’ with multiple energy minima. A number of experimental results support the presence of metastable states in manganites. For example, magnetic fields²⁵ or X-rays²⁶ have been used to convert insulating regions into ferromagnetic metallic ones, which are stable even when the magnetic fields or X-rays are removed. For illustration, we consider here two cases with different harmonic moduli for short-wavelength lattice distortions as shown in Fig. 1b, one giving a shallow (blue curve, a small modulus) and the other a deep (red curve, a large modulus) local minimum for the undistorted phase. Solid circles in the inset represent locations of energy minima for our model in the $s_x - s_y$ space. The modulus can be changed, for example, by varying the average size of the rare-earth (Re) and alkali-metal (Ak) ions, which sensitively alters manganite properties¹⁵.

Figure 2a–h shows the sequential relaxations for the shallow local minimum case starting from random initial s_x and s_y values, corresponding to a rapid cooling of the system from high tempera-

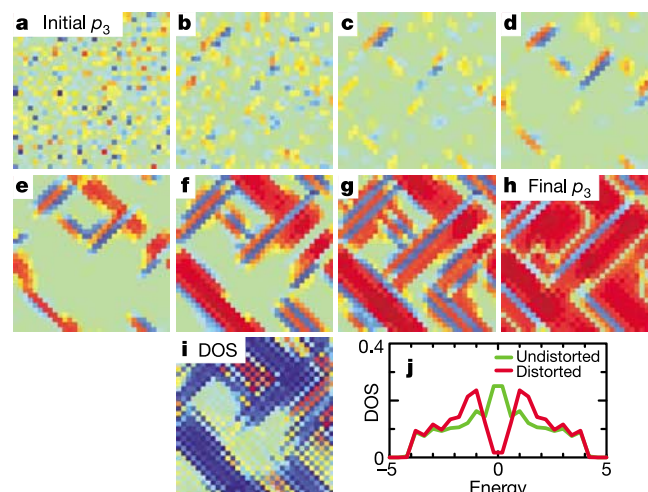


Figure 2 Results of simulations for the shallow local minimum case (blue curve in Fig. 1b) on a 32×32 lattice with periodic boundary conditions. **a**, The $p_3 = s_x^2 - s_y^2$ field initial configuration. **b–g**, Time-sequence of p_3 during energy relaxation. **h**, Final stable stage. (Dark red and dark blue represent $\pm s_0^2$, except in **a** where they correspond to $\pm (2.6s_0)^2$. s_0 is the value of s_x or s_y at the global minimum distorted state.) **i**, Local electronic DOS at $\epsilon_F = 0$ for the distorted pattern in **f** (dark blue and dark red correspond to 0 and 0.5). **j**, Typical local electronic DOS versus electronic energy within the distorted and undistorted regions. The small finite DOS within the gap is due to the exponentially decaying leakage of electronic state from the metallic region: a ‘pseudogap’. The distorted lattice with s_x and negative e_3 , or equivalently that with s_y and positive e_3 (see Fig. 1a), leads to a gap in the electronic DOS near electronic energy $\epsilon = 0$. Thus, if ϵ_F lies in the gap, the distorted lattice behaves as an insulator. In the structure without distortions, the DOS has no gap, and the electrons are in a metallic state. Crucially, the study of the inhomogeneous phase shown here in **a–h** requires that constraints between the distortion variables be satisfied, allowing bonds between atoms to bend but not break^{12,13,29}. Elimination of e_1, e_2 and e_3 subject to these constraint equations leads to an energy expression in terms of s_x and s_y , which we numerically minimize using various initial configurations. The electronic properties in **i** and **j** are obtained by numerically solving the SSH hamiltonian on the elastic templates.

ture. The results are represented in terms of $p_3 = s_x^2 - s_y^2$, which serves as an order parameter. Positive (red) and negative (blue) values of p_3 correspond to different orientations of short- or long-wavelength lattice distortions (s_x with negative e_3 and s_y with positive e_3). The short- and long-wavelength mode distortions are simultaneously generated through the minimal symmetry-allowed coupling between e_3 and $s_x^2 - s_y^2$ at each site. The green regions with zero p_3 have no distortions. Most of the region initially relaxes to the undistorted local minimum state, as shown in Fig. 2c and d, because the rapidly fluctuating initial field contains few components of the long-wavelength strain modes, which makes it difficult for the system to reach the global minimum state. However, even in the presence of the random field, there are regions with some correlation, which eventually lead to nanometre-scale nucleating droplets. Comparison of Fig. 2b–d shows that there is a critical strength and size of these droplets, a common feature of first-order phase transitions. However, the long-range nature and anisotropy of the interaction between strain fields^{12,13,29} define the nucleation process uniquely. First, the morphology of the nucleated droplets often consists of a pair of strain fields with different orientations (see Fig. 2d) to minimize the energy cost at the boundary between the distorted and undistorted phases. Second, extended long-range correlations along the favoured direction can allow for easy growth of distorted regions (the broad 135-degree $p_3 > 0$ regions in Fig. 2d and e). Third, the nucleated droplets are highly anisotropic, unlike usual first-order phase transitions. These distinct characteristics of elastic nucleation are the probable source of the nanometre-scale

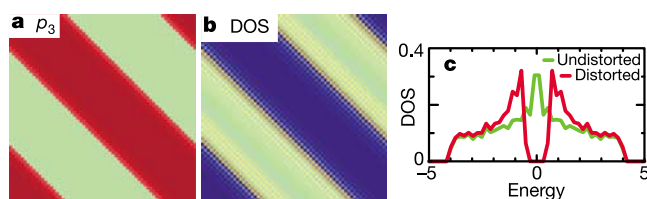


Figure 3 Results of simulations for the deep local minimum case (red curve in Fig. 1b) on a 64×64 lattice. **a**, p_3 field for the stable elastic texture (same colour scheme as Fig. 2). **b**, Local DOS configuration at $\epsilon_F = 0$. Blue and green regions are distorted insulator and undistorted metal, respectively. **c**, Typical local DOS versus energy within distorted and undistorted regions. As seen by comparing Figs 2 and 3, the depth of the local energy minimum changes the nature of the inhomogeneity, from nanometre-scale fluctuations to micrometre-scale stable domains. In manganites, the harmonic modulus for the short-wavelength distortions, which changes the depth of the local energy minimum in Fig. 1b, may be varied by changing the size of Re/Ak ions (often parameterized as a 'tolerance factor'¹⁵). More specifically, the effect of the size of Re/Ak ions³⁰ on the modulus may be represented through an additional symmetry-allowed coupling $C_1 e_1 (s_x^2 + s_y^2)$ in our model. The small-size Re/Ak site ions create an isotropic strain (that is, e_1), which renormalizes the harmonic modulus for the short-wavelength distortion modes through this cubic anharmonic term. Other physics, such as magnetism and kinetic energy of electrons, particularly coupled with the buckling of oxygen octahedra, can also change the effective modulus and the energy landscape.

inhomogeneity seen in manganites, for example, the anisotropic short-range precursor correlations at high temperatures⁸ and the quasielastic central peak in neutron scattering near the metal–insulator phase-transition temperatures²² (analogues of the central peaks in ferroelectrics). Such behaviour is very reminiscent of the precursor embryonic fluctuations observed in martensitic transformations²⁰.

Figure 2i and j shows the tight-binding electronic structure supported by the lattice distortion template in Fig. 2f. For illustration, we have used a Su–Shrieffer–Heeger (SSH) electron–lattice coupling model¹³ (with one electron per site) that linearly incorporates the variation of the electron transfer probability between neighbouring sites due to the changes in interatomic distances. This SSH model, together with the specific coupling between short- and long-wavelength modes mentioned earlier, gives rise to the desired electronic phases for undistorted and distorted structures. The typical local electronic density of states (DOS) versus electronic energy shown in Fig. 2j indeed indicates a metallic local DOS in the undistorted region and a gapped (insulating) local DOS within the distorted region. For the Fermi energy $\epsilon_F = 0$, the local DOS configuration in direct space is shown in Fig. 2i. This illustrates the coexistence of metal (green) and insulator (blue) associated with the elastic texture template, similar to that observed in manganites^{2–4}.

For the deeper local-minimum case, the simulation of the rapid cooling process from random initial configurations shows characteristic features of metastability or a supercooled state, such as a low probability of creating nucleation droplets. Such slow relaxation dynamics requires extensive computational resources, so we choose instead predesigned initial conditions, such as a sinusoidal spatial variation of s_x or s_y , along a 45-degree direction. A typical result in the final p_3 field is shown in Fig. 3a. The regions with large (or small) initial $|s_x|$ or $|s_y|$ transform to a distorted global (or undistorted local) minimum state upon energy minimization. The energy landscape near the undistorted state is deep enough that the coexisting phase of metal and insulator is stable, unlike the shallow local-minimum case above. Such stable coexistence of metallic and insulating domains, which can be as large as several micrometres when simulated on a larger lattice, is similar to the submicrometre size multiphase coexistence observed in $\text{La}_{1-x-y}\text{Pr}_y\text{Ca}_x\text{MnO}_3$ (ref. 2). Thermal fluctuations at finite temperatures can grow the distorted

region slowly—such slow, history-dependent dynamics with time-scales that range from 10–20 min to 2–3 h has been also observed in various manganites^{22,24}. Figure 3b and c shows the corresponding electronic inhomogeneities within the SSH model. The local DOS for the sites deep within the domain shows a clear gap near electronic energy $\epsilon = 0$. Thus, for $\epsilon_F = 0$, the metallic and insulating phases coexist with relatively sharp boundaries, as observed in STM images of $\text{Bi}_{0.24}\text{Ca}_{0.76}\text{MnO}_3$ (ref. 4).

A magnetic field is expected to modify the energy landscape in favour of the undistorted ferromagnetic metallic state, which would increase the volume fraction of metallic regions in both the phase with metal–insulator domains (Fig. 3) and the phase with precursor embryonic fluctuations (Fig. 2). In particular, the reduction of the quasi-elastic scattering spectral weight by applied magnetic fields²³ provides evidence for the latter case, and hence percolating conducting paths created by these fields can lead to the observed colossal magnetoresistance. The mechanisms we have proposed here can be applied to describe inhomogeneities in other materials with strong bonding constraints, such as relaxor ferroelectrics and possibly high-transition-temperature (high- T_c) superconducting oxides, where we believe the functionalities are also mediated through self-organized lattice distortions.

The above results show that the micrometre-scale multiphase coexistence is self-organized and is caused by the presence of an intrinsic elastic energy landscape, whereas existing theory^{9–11} suggests that the micrometre-scale inhomogeneity originates from the random distribution of Re/Ak ions. The two models make different predictions on engineering patterns of metal and insulator domains. The existing theory requires redistribution of Re/Ak ions to change patterns of multiphase domains. In our model, because the domain formation is self-sustained, external stimuli such as optical lasers, X-rays, or ultrasonic standing waves, can be used to sensitively manipulate patterns of metallic and insulating regions, thus making the control of nano-engineered functional domains in manganites feasible. □

Received 4 November 2003; accepted 23 January 2004; doi:10.1038/nature02364.

- Salamon, M. B. & Jaime, M. The physics of manganites: structure and transport. *Rev. Mod. Phys.* **73**, 583–628 (2001).
- Uehara, M., Mori, S., Chen, C. H. & Cheong, S.-W. Percolative phase separation underlies colossal magnetoresistance in mixed-valent manganites. *Nature* **399**, 560–563 (1999).
- Fäth, M. *et al.* Spatially inhomogeneous metal–insulator transition in doped manganites. *Science* **285**, 1540–1542 (1999).
- Renner, Ch., Aepli, G., Kim, B.-G., Soh, Y.-A. & Cheong, S.-W. Atomic-scale images of charge ordering in a mixed-valence manganite. *Nature* **416**, 518–521 (2002).
- Zhang, L., Israel, C., Biswas, A., Greene, R. L. & de Lozanne, A. Direct observation of percolation in a manganite thin film. *Science* **298**, 805–807 (2002).
- Mathur, N. & Littlewood, P. Mesoscopic texture in manganites. *Phys. Today* **56**, 25–30 (2003).
- Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La–Ca–Mn–O films. *Science* **264**, 413–415 (1994).
- Kim, K. H., Uehara, M. & Cheong, S.-W. High-temperature charge-ordering fluctuation in manganites. *Phys. Rev. B* **62**, R11945–R11948 (2000).
- Moreo, A., Yunoki, S. & Dagotto, E. Phase separation scenario for manganese oxides and related materials. *Science* **283**, 2034–2039 (1999).
- Dagotto, E., Hotta, T. & Moreo, A. Colossal magnetoresistant materials: The key role of phase separation. *Phys. Rev. B* **344**, 1–153 (2001).
- Burg, J., Moreo, A. & Dagotto, E. Relevance of cooperative lattice effects and correlated disorder in phase-separation theories for CMR manganites. *Phys. Rev. Lett.* (submitted); preprint at (<http://www.arXiv.org/cond-mat/0308456>) (2003).
- Ahn, K. H., Lookman, T., Saxena, A. & Bishop, A. R. Atomic scale lattice distortions and domain wall profiles. *Phys. Rev. B* **68**, 092101 (2003).
- Ahn, K. H., Lookman, T., Saxena, A. & Bishop, A. R. Microstructural evolution and electronic properties of antiphase boundaries in elastic materials. *Phys. Rev. B* (submitted); preprint at (<http://www.arXiv.org/cond-mat/0309328>) (2003).
- Millis, A. J. Lattice effects in magnetoresistive manganese perovskites. *Nature* **392**, 147–150 (1998).
- Hwang, H. Y., Cheong, S.-W., Radaelli, P. G., Marezio, M. & Batlogg, B. Lattice effects on the magnetoresistance in doped LaMnO_3 . *Phys. Rev. Lett.* **75**, 914–917 (1995).
- Mathur, N. D. & Littlewood, P. B. The self-organised phases of manganites. *Solid State Commun.* **119**, 271–280 (2001).
- Millis, A. J. Towards a classification of the effects of disorder on materials properties. *Solid State Commun.* **126**, 3–8 (2003).
- Bishop, A. R., Lookman, T., Saxena, A. & Shenoy, S. R. Elasticity-driven nanoscale texturing in complex electronic materials. *Europhys. Lett.* **63**, 289–295 (2003).
- Podzorov, V., Kim, B. G., Kiryukhin, V., Gershenson, M. E. & Cheong, S.-W. Martensitic accommodation strain and the metal–insulator transition in manganites. *Phys. Rev. B* **64**, 140406 (2001).

20. Seto, H., Noda, Y. & Yamada, Y. Precursor phenomena at martensitic phase transition in Fe-Pd alloy. II. Diffuse scattering and embryonic fluctuations. *J. Phys. Soc. Jpn* **59**, 978–986 (1990).
21. Ferrari, V., Towler, M. & Littlewood, P. B. Oxygen stripes in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ from ab initio calculations. *Phys. Rev. Lett.* **91**, 227202 (2003).
22. Lynn, J. W. *et al.* Unconventional ferromagnetic transition in $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. Lett.* **76**, 4046–4049 (1996).
23. Lynn, J. W. *et al.* Magnetic, structural, and spin dynamical properties of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *J. Appl. Phys.* **81**, 5488–5490 (1997).
24. Levy, P., Parisi, F., Granja, L., Indelicato, E. & Polla, G. Novel dynamical effects and persistent memory in phase separated manganites. *Phys. Rev. Lett.* **89**, 137001 (2002).
25. Tokura, Y., Kuwahara, H., Moritomo, Y., Tomioka, Y. & Asamitsu, A. Competing instabilities and metastable states in $(\text{Nd,Sm})_{1/2}\text{Sr}_{1/2}\text{MnO}_3$. *Phys. Rev. Lett.* **76**, 3184–3187 (1996).
26. Kiryukhin, V. *et al.* An X-ray induced insulator metal transition in a magnetoresistive manganite. *Nature* **386**, 813–815 (1997).
27. Fiebig, M., Miyano, K., Tomioka, Y. & Tokura, Y. Visualization of the local insulator-metal transition in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Science* **280**, 1925–1928 (1998).
28. Chen, C. H. & Cheong, S.-W. Commensurate to incommensurate charge ordering and its real-space images in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Phys. Rev. Lett.* **76**, 4042–4045 (1996).
29. Lookman, T., Shenoy, S. R., Rasmussen, K. Ø., Saxena, A. & Bishop, A. R. Ferroelastic dynamics and strain compatibility. *Phys. Rev. B* **67**, 024114 (2003).
30. Rodriguez-Martinez, L. M. & Attfield, J. P. Cation disorder and size effects in magnetoresistive manganese oxide perovskites. *Phys. Rev. B* **54**, R15622–R15625 (1996).

Acknowledgements We thank A. Saxena for discussions. The work was supported by the US DOE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.H.A. (ahn@lanl.gov).

Onset of heterogeneous crystal nucleation in colloidal suspensions

A. Cacciuto, S. Auer* & D. Frenkel

FOM Institute for Atomic and Molecular Physics, Kruislaan 407,
1098 SJ Amsterdam, The Netherlands

* Present address: Department of Chemistry, Cambridge University, Lensfield Road, Cambridge,
CB2 1EW, UK

The addition of small ‘seed’ particles to a supersaturated solution can greatly increase the rate at which crystals nucleate. This process is understood, at least qualitatively, when the seed has the same structure as the crystal that it spawns^{1,2}. However, the microscopic mechanism of seeding by a ‘foreign’ substance is not well understood. Here we report numerical simulations of colloidal crystallization seeded by foreign objects. We perform Monte Carlo simulations to study how smooth spherical seeds of various sizes affect crystallization in a suspension of hard colloidal particles. We compute the free-energy barrier associated with crystal nucleation^{3,4}. A low barrier implies that nucleation is easy. We find that to be effective crystallization promoters,

the seed particles need to exceed a well-defined minimum size. Just above this size, seed particles act as crystallization ‘catalysts’ as newly formed crystallites detach from the seed. In contrast, larger seed particles remain covered by the crystallites that they spawn. This phenomenon should be experimentally observable and can have important consequences for the control of the resulting crystal size distribution.

Ostwald’s 1897 paper on crystal nucleation¹ contains an interesting remark in the introduction. After explaining that supercooled liquids can be made to crystallize by the introduction of a small seed crystal, Ostwald writes: “I am not aware of any experiments to determine the smallest amount of solid that is needed to make this procedure [that is, the crystallization] succeed”.

More than a century has passed since then, and much has been learned about the process of crystal nucleation. But part of the question that Ostwald posed remains unanswered. The classical theory of nucleation provides a natural explanation of why a seed crystal facilitates crystal nucleation²: in order to grow, crystallites of the stable phase need to exceed a critical size. Crystallites that are smaller than this ‘critical nucleus’ dissolve again, crystallites that are larger can grow to a macroscopic size. In the absence of a seed, a rare, spontaneous fluctuation is needed to form a crystal nucleus that exceeds the critical size. However, crystallization can proceed spontaneously if we add a seed crystal that is larger than the critical nucleus to the metastable liquid phase.

But crystallization can also be induced by introducing foreign substances. And for that case, the answer to Ostwald’s question is not known. Here we use computer simulations to study how very small foreign objects (‘nano-dirt’) influence the rate of crystal nucleation. In particular, we consider the effect of spherical seed particles on the crystallization of uncharged, spherical colloids.

First, we consider the effect of a flat wall on crystal nucleation. On an atomic scale, no wall is perfectly flat. But it is relatively straightforward to study colloidal crystallization on surfaces that are flat on the scale of a colloidal particle (50 nm, or more). Experiments^{5–7} and simulations⁸ show that such a flat wall speeds up colloidal crystallization by many orders of magnitude. Next, consider a smooth, spherical object of finite size. Clearly, if the object is very large, its effect should be similar to that of a hard wall. At the other limit, where the object has the same size as the colloids in the system, the object should have no effect on the nucleation rate.

One might expect that nucleation speeds up monotonically with the size of the spherical seed, but this is not true. Because crystals cannot grow on spheres without generating topological defects^{9,10}, nucleation of spherical particles on, or near, a sphere with a different size is inhibited. This shows up dramatically in the increase of the crystal nucleation barrier in a system of polydisperse hard spheres¹¹. Yet, if we introduce only a single, mismatched sphere in the liquid, then the effect on nucleation is minimal: true, nucleation does not take place very close to this sphere, but the homogeneous nucleation in the rest of the volume is unaffected. To discover the minimum size of a spherical seed that can affect the rate of crystal nucleation,

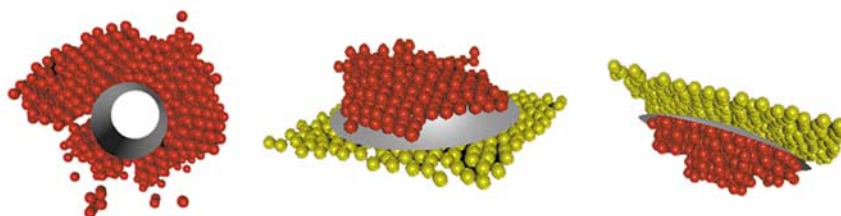


Figure 1 The spontaneous crystallizations of hard-sphere colloids on seed particles. The seed particles are a cylinder ($R_s = 3\sigma$, at $\phi = 0.523$) and a spherical cap ($R_s = 60\sigma$, at $\phi = 0.513$). The middle and right images show nucleation on the convex and on the

concave sides of a spherical seed, respectively. To distinguish between the two cases, we cover the bottom (for $\kappa > 0$) or top (for $\kappa < 0$) of the spherical caps with a disordered template to prevent crystallization on that side of the seed.

we performed Monte Carlo numerical simulations of a system of monodisperse hard colloids with diameter σ containing a seed of radius R_s at its centre and systematically explored the effect of a wide range of seed sizes.

Spherical seeds with large radii ($R_s > 10\sigma$) can be studied in principle, but the computational effort quickly becomes prohibitive. To alleviate this problem, we performed simulations in which we used, as a seed, only part of the surface of a sphere with a large radius of curvature.

As a first test, we studied nucleation on a flat, circular template (infinite radius limit) with surface area $A = 150\sigma^2$ in a cubic box containing 5,632 hard spheres. The volume fraction of colloids in the liquid at freezing is $\phi_{\text{coex}}^1 = 0.494$. In our simulations, we quickly compressed the fluid to a volume fraction $\phi = 0.513$ (reduced pressure $P\sigma^3/k_B T = 14$ where k_B is the Boltzmann constant and T is absolute temperature). At this pressure, the rate of homogeneous crystal nucleation is vanishingly small, yet, as expected⁸, crystallization occurred immediately on the circular template.

Subsequently, we studied crystallization on a series of seeds, all with the same area, but with radii ranging from $R_s = 150\sigma$ to $R_s = 10\sigma$. We detected the onset of crystallization on both the convex and the concave sides of these seeds (Fig. 1). As can be seen in Fig. 2, higher pressures are needed to induce crystallization on the more strongly curved templates. Figure 2 shows that crystallization on a spherical template takes place more readily on the convex than on the concave side. The excess pressure ΔP , that is, the pressure at which crystallization sets in on a curved surface minus the corresponding pressure for a flat surface⁸, appears to scale as $\Delta P \approx \kappa^\alpha$, where $\kappa \equiv R_s^{-1}$. A fit to the simulation data yielded $\alpha = 0.7(1)$.

We also studied seeded crystal nucleation on cylindrical templates and observed a similar scaling relation between the excess pressure at crystallization and the radius of curvature (see Fig. 2). However, for cylinders crystallization consistently takes place at lower pressures than for spheres. The observed behaviour can be described, at least qualitatively, by an extension of classical nucleation theory that takes into account the free-energy cost of elastic deformations (A.C. and D.F., manuscript in preparation).

At a pressure $P\sigma^3/k_B T = 16$ ($\phi = 0.5277$), we could study crystal nucleation on a complete spherical seed. At this pressure, the barrier for homogeneous nucleation ($\Delta G \approx 28k_B T$) is still high enough to rule out spontaneous crystallization in the bulk, and the size of the critical nucleus, $n_c \approx 110$, is sufficiently small to neglect finite size effects. We considered a system consisting of 12,167 particles and we introduced complete spherical seeds of radius $R_s = 4\sigma, 5\sigma, 6\sigma$ and 7σ .

Using the umbrella-sampling techniques that we have developed^{3,4}, we estimated ΔG , the height of the crystal nucleation

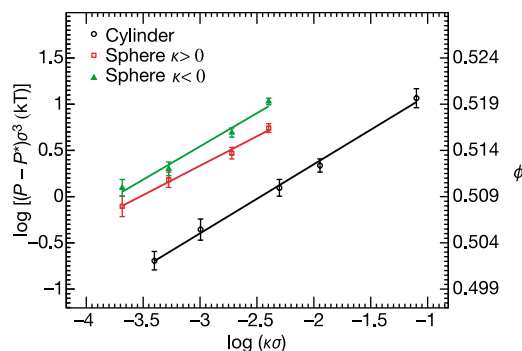


Figure 2 Plot of the smallest pressure P that causes complete crystallization as a function of the seed curvature κ . (Minus $P^* = 12.2$, representing the 'flat wall' limit.) The curves refer to the cylinder and the convex ($\kappa > 0$) and concave ($\kappa < 0$) sides of the spherical seed, as indicated. On the right axis, we show the corresponding volume fractions ϕ .

barrier, for the various spherical seeds (see Fig. 3). In the same figure, we also show the barrier for homogeneous nucleation. For a seed with radius $R_s = 4\sigma$, we find that crystal nuclei typically form in the bulk, rather than on the seed. Hence, spheres of this size are not crystallization promoters.

Seeded nucleation starts with a spherical shell of radius $R_s \approx 5\sigma$: very small nuclei appear on the seed surface. These tend to grow radially outward until they detach and move away into the bulk (see Fig. 4). Although this seed clearly affects the crystal nucleation pathway, it hardly affects the height of the nucleation barrier. As can be seen from Fig. 3, larger seeds do lower this barrier. However, the size of the critical nucleus is hardly sensitive to the radius of the seed. This may be understood by inspecting the pathway for nucleation followed in this case (see Fig. 4).

Figure 4 shows that small nuclei grow on the seed. As they grow, the presence of a curved substrate makes it difficult to maintain an unstrained structure. At some stage, the precritical nuclei break away from the surface, and the critical nucleus is only formed in the bulk. This nucleus is therefore similar to the one observed in homogeneous nucleation. As each crystal nucleus detaches from it, the seed again becomes free to produce a new crystal—it can therefore act as an 'assembly line' for crystal nuclei (in the same way that boiling chips produce many small bubbles). Note that this will not happen when the seed is a crystal of the same material. This has important consequences for the size distribution of crystallites formed in heterogeneous nucleation: large, nearly flat seeds tend to produce one single crystal that grows to macroscopic size. But intermediate size, foreign seeds keep on producing crystal nuclei. As a consequence, the number of crystallites will be much larger than

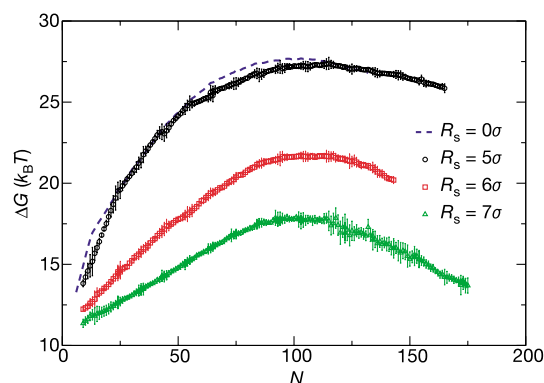


Figure 3 Free-energy barriers for crystal nucleation in a system of hard spheres with a smooth spherical seed. The seed has radii $R_s = 5\sigma, 6\sigma$ and 7σ . The dashed curve represents the homogeneous nucleation barrier.

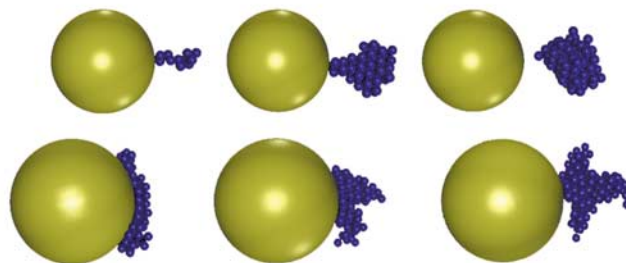


Figure 4 Snapshot sequence of crystal nucleation on spherical seeds. The seeds have radii $R_s = 5\sigma$ (top) and 7σ (bottom). From left to right the sizes of the nuclei are, respectively, $N \approx 10, \approx 50$ and ≈ 100 for the sphere of radius $R_s = 5\sigma$ and $N \approx 60, \approx 80$ and ≈ 120 for the sphere of radius $R_s = 7\sigma$. The above phenomenon should be observable in confocal microscopy studies of seeded crystallization in a colloidal suspension.

the number of seeds and the result is that the final crystallites will be small. In many industrial processes, the control of crystallite size is important. The present work suggests that a careful selection of the size of the crystallization promoter can be of crucial importance.

Real-space imaging techniques, such as the ones used to study crystallization in colloidal suspensions¹², could be used to study seeded crystal nucleation in real colloidal systems. In such experiments, relatively large colloidal beads (spheres) or glass fibres (cylinders) could be introduced in a supersaturated solution of monodisperse, 'hard sphere' colloids. Crystallization should take place readily on the larger seeds. Owing to gravity, crystal nuclei that detach from the seed should sediment away from it. This should make it easier to image subsequent crystal nucleation events on the same seed. □

Received 9 September 2003; accepted 3 February 2004; doi:10.1038/nature02397.

- Ostwald, W. F. Studien über die Bildung und Umwandlung fester Körper. *Z. Phys. Chem.* **22**, 289–302 (1897).
- Kelton, K. E. In *Solid State Physics* (eds Ehrenreich, H. & Turnbull, D.) Vol. 45 (Academic, New York, 1991).
- ten Wolde, P. R., Ruiz-Montero, M. J. & Frenkel, D. Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *J. Chem. Phys.* **104**, 9932–9947 (1996).
- Auer, S. & Frenkel, D. Quantitative prediction of crystal nucleation rates for spherical colloids: A computational approach. *Annu. Rev. Phys. Chem.* **55**, (2004).
- Kose, A. & Hachisu, S. Ordered structure in weakly flocculated monodisperse latex. *J. Colloid Interf. Sci.* **55**, 487–498 (1976).
- Dinsmore, A., Warren, P., Poon, W. & Yodh, A. Fluid-solid transitions on walls in binary hard-sphere mixtures. *Europhys. Lett.* **40**, 337–342 (1997).
- Kaplan, P., Rouke, J., Yodh, A. & Pine, D. Entropically driven surface phase separation in binary colloidal mixtures. *Phys. Rev. Lett.* **72**, 582–585 (1994).
- Auer, S. & Frenkel, D. Line tension controls wall-induced crystal nucleation in hard-sphere colloids. *Phys. Rev. Lett.* **91**, 015703 (2003).
- Euler, L. *Opera Omnia* Series i, Vol. 26 (Orell Füssli, Zurich, 1953).
- Hilton, P. & Pederson, J. The Euler characteristic and Pólya's dream. *Am. Math. Mon.* **103**, 121–131 (1996).
- Auer, S. & Frenkel, D. Suppression of crystal nucleation in polydisperse colloids due to increase of the surface free energy. *Nature* **413**, 711–713 (2001).
- Gasser, U., Weeks, E. R., Schofield, A., Pussey, P. N. & Weitz, D. A. Real-space imaging of nucleation and growth in colloidal crystallization. *Science* **292**, 258–262 (2001).

Acknowledgements The work of the FOM Institute is part of the research program of FOM and is made possible by financial support from the Netherlands Organization for Scientific Research (NWO).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C. (cacciuto@amolf.nl).

Mass and volume contributions to twentieth-century global sea level rise

Laury Miller¹ & Bruce C. Douglas²

¹Laboratory for Satellite Altimetry, NESDIS, NOAA, Silver Spring, Maryland 20910, USA

²Laboratory for Coastal Research, Florida International University, Miami, Florida 33199, USA

The rate of twentieth-century global sea level rise and its causes are the subjects of intense controversy^{1–7}. Most direct estimates from tide gauges give 1.5–2.0 mm yr⁻¹, whereas indirect estimates based on the two processes responsible for global sea level rise, namely mass and volume change, fall far below this range. Estimates of the volume increase due to ocean warming give a rate of about 0.5 mm yr⁻¹ (ref. 8) and the rate due to mass

increase, primarily from the melting of continental ice, is thought to be even smaller. Therefore, either the tide gauge estimates are too high, as has been suggested recently⁶, or one (or both) of the mass and volume estimates is too low. Here we present an analysis of sea level measurements at tide gauges combined with observations of temperature and salinity in the Pacific and Atlantic oceans close to the gauges. We find that gauge-determined rates of sea level rise, which encompass both mass and volume changes, are two to three times higher than the rates due to volume change derived from temperature and salinity data. Our analysis supports earlier studies that put the twentieth-century rate in the 1.5–2.0 mm yr⁻¹ range, but more importantly it suggests that mass increase plays a larger role than ocean warming in twentieth-century global sea level rise.

At the time of the second IPCC assessment in 1995⁹, there seemed to be little controversy regarding global sea level rise (GSLR). Most gauge estimates fell in the range 1.5–2.0 mm yr⁻¹. Most of this rise was thought to result from ocean warming, with the rest due to melting of continental ice. However, by the third IPCC assessment in 2001¹, this consensus view had collapsed: new and better estimates of ocean warming had reduced the volume increase component to about 0.5 mm yr⁻¹ (ref. 8), and the mass component was thought to be even smaller. This left a large unexplained gap between direct and indirect estimates of GSLR, now known as the 'attribution problem'.

Two recent studies offer opposing solutions to this dilemma. Cabanes *et al.*⁶ argue that gauge rates are 2–3 times too high because the gauges happen to be located in areas of abnormally high ocean warming. They arrive at this result by comparing gauge-derived sea level trends with those obtained from objectively interpolated hydrographic measurements, concluding that the true rate of GSLR is actually 0.5–1.0 mm yr⁻¹, mostly due to ocean warming. This solution provides a way out of the attribution problem, but implies a huge acceleration of GSLR in the 1990s if recent satellite altimetric estimates of ~2.5 mm yr⁻¹ (ref. 10) are to be believed. Alternatively, Antonov *et al.*⁷ suggest that the problem may be solved by revising upward the mass component estimate. They show that the oceans are freshening at a rate equivalent to the addition of 1.4 mm yr⁻¹ of fresh water, approximately the value needed to bring the mass plus volume rate close to the gauge rate. However, this solution assumes a continental ice source rather than floating ice, a key point that they are unable to demonstrate.

Here we present a simple approach to the problem of distinguishing between mass and volume contributions to GSLR. We identify large areas in the Pacific and Atlantic oceans that are either bounded by, or adjacent to, several gauge sites exhibiting similar trends and variability. For those regions, we compare average gauge trends (which reflect both mass and volume change) with average dynamic height trends (which reflect only volume change), with the added distinction that we use raw rather than interpolated hydrographic data. In most areas, the difference between raw and interpolated data are unimportant. However, as shown below, the interpolated data used by Cabanes *et al.*⁶ contains errors near the Gulf Stream large enough to cause their average warming estimate for the gauge sites to be biased upward by a factor of 2–3. In contrast, we find no evidence that the gauges are located in regions of abnormally high warming.

We begin in the Eastern Pacific in a region bounded by gauge sites at Honolulu, San Francisco, San Diego, and Balboa, Panama (map inset, Fig. 1). This region has several characteristics that make it ideal for this study: low mesoscale variability, narrow continental margins, large numbers of hydrographic observations, and long (>90 yr) gauge records. Figure 1 presents ~19,000 dynamic height anomaly observations relative to 1,000 m, their 5-yr running means, and 5-yr running means of the four gauge records. Although the hydrographic data appear noisy in this plot, their r.m.s., 6.7 cm, is comparable to satellite altimeter determinations of sea level variability in this relatively quiet part of the global ocean¹¹. A linear

regression on the smoothed dynamic heights indicates an upward trend of 0.5 mm yr^{-1} between 1932 and 1997. Similar trends are found when the calculations are performed on the much smaller numbers of observations reaching 2,000 m and 3,000 m (3,174 and 2,119, respectively).

In contrast to the hydrographic results, all four gauge records show trends of $\sim 2 \text{ mm yr}^{-1}$ during the twentieth century, despite the fact that the gauge sites are widely separated and thus subject to different vertical land motions and local hydrographic conditions. Additional evidence that the gauges are observing the same regional trend comes from the fact that all four records also show similar interannual variations dominated by El Niño events.

Could the discrepancy between the gauge-derived and hydrographic-measurement-derived trends in Fig. 1 be explained by unusually high rates of warming near the gauge sites? Figure 2 shows sea level and dynamic height comparisons for three local areas encompassing (or adjacent to) the gauge sites. Because of strong similarities between the San Francisco and San Diego gauge records, the California coast is treated as one region in Fig. 2a. California and Honolulu (Fig. 2a, b) yield 1,000-m dynamic height trends of 0.5 mm yr^{-1} and 0.3 mm yr^{-1} from 1938 to 1996 and 1950 to 1999, respectively. Extending the hydrographic analysis in either of these locations to 2,000 m does not significantly change the results (there are insufficient hydrographic data in either region to go to 3,000 m). Balboa (Fig. 2c) exhibits a larger 1,000-m dynamic height trend, 0.9 mm yr^{-1} , but still only half the matching gauge trend of 1.7 mm yr^{-1} over the interval 1930–1993. Results from each local area in the eastern Pacific therefore support the conclusion summarized in Fig. 1: that twentieth-century sea level rose at a rate several times higher than can be accounted for by volume (temperature and salinity) changes alone.

Figure 3 shows a comparison of dynamic height and sea level trends in the eastern North Atlantic, another area with low meso-scale variability, narrow margins, and so on. The dynamic height trend there is only 0.2 mm yr^{-1} from 1911 to 1998. In contrast, the gauges at Cascais (Portugal) and Tenerife (Canary Islands) show sea level trends of 2.1 and 1.9 mm yr^{-1} , respectively, an order of magnitude greater than the hydrographic data. Extending the hydrographic analysis more deeply into the water column does not significantly alter this result.

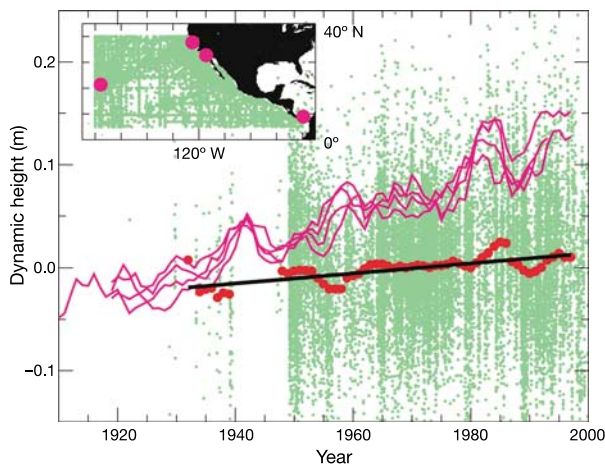


Figure 1 Regional comparison of hydrographic and tide-gauge-measured sea level change in the Eastern Pacific. Hydrographic profile observations of temperature and salinity converted into 1,000-m dynamic height anomalies (green), their 5-yr running means (red data points) and linear regression (black) compared with 5-yr running mean relative sea levels from tide gauge observations at San Francisco, San Diego, Honolulu and Balboa (purple lines). Tide gauge series have been vertically offset to coincide with the earliest dynamic heights. Map inset shows tide gauge locations in purple and observed dynamic height locations in green.

Finally, Fig. 4 presents an analysis of the Slope Water region in the western North Atlantic, north of the Gulf Stream and east of the tide gauge sites at Boston, Portland and Halifax. Two types of dynamic height data are shown. The green dots, and their 5-yr means in red, represent dynamic height anomalies computed from observed temperature and salinity measurements, as in Figs 1–3. The blue curves represent anomaly values computed from the World Ocean Atlas 1998v2 (WOA98v2)¹², the objectively interpolated hydrographic data set used by Cabanes *et al.*⁶.

As in the eastern Pacific and Atlantic, the gauge records exhibit substantially higher trends, an average 1.9 mm yr^{-1} during the twentieth century, than those based on the observed hydrographic data. In fact, the trends on the latter are negligible over the record (1931–1996). By contrast, the results based on the WOA98v2 analysis exhibit an abrupt 20-cm increase between 1965 and 1975 that is not present in either the surrounding hydrographic observations or gauge records. Over the 1955–1996 interval used by Cabanes *et al.*⁶, the average WOA98v2-derived trend is 6.7 mm yr^{-1} , more than three times greater than the corresponding gauge trend.

There is good reason to believe that this ‘jump’ is an artefact of the large (444 to 888 km) variable radius of influence used in the

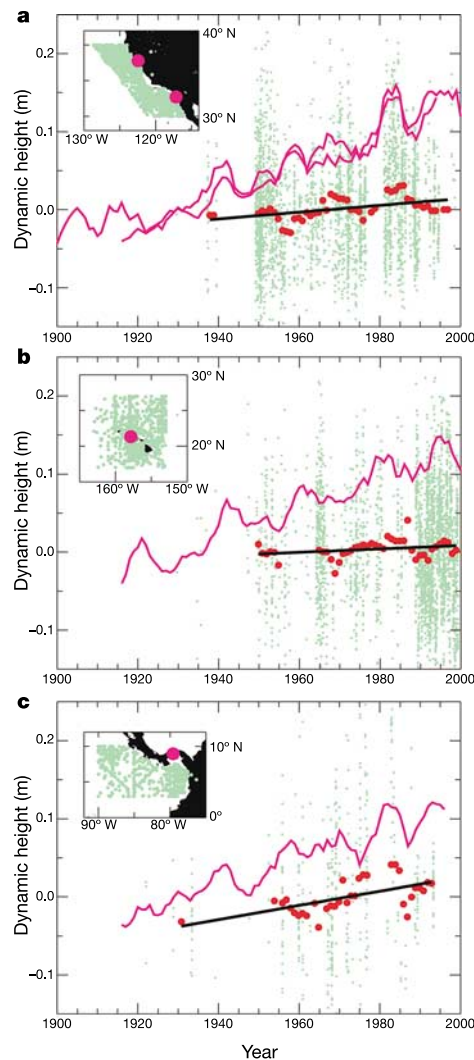


Figure 2 Local comparison of hydrographic and tide-gauge-measured sea level change. **a**, As in Fig. 1, but hydrographic observations limited to $4,400 \text{ km} \times 1,100 \text{ km}$ area adjacent to gauge sites at San Francisco and San Diego; **b**, to $1,100 \text{ km} \times 1,100 \text{ km}$ area centred on gauge site at Honolulu; **c**, to $800 \text{ km} \times 1,400 \text{ km}$ area adjacent to gauge site at Balboa, Panama.

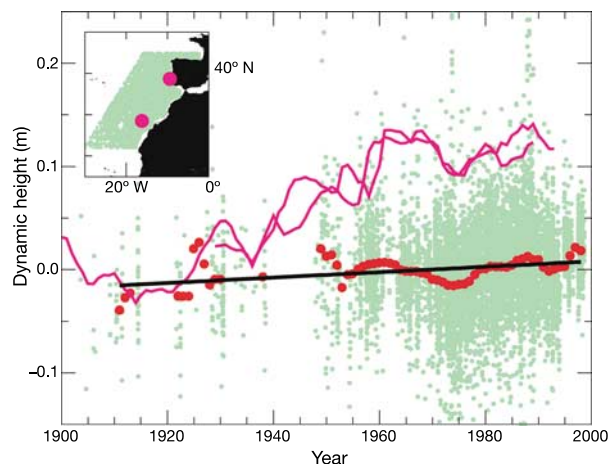


Figure 3 Regional comparison of hydrographic and tide-gauge-measured sea level change in the Eastern Atlantic. As in Fig. 1, but hydrographic observations limited to 1,400 km $\times$ 2,200 km area adjacent to tide gauge sites at Cascais (Portugal) and Tenerife (Canary Islands).

WOA98v2 objective analysis. Between the mid-1960s and early 1970s, the mean position of the Gulf Stream shifted northward by about 50 km as a result of gyre-scale changes in the surface wind field¹³. An examination of the actual hydrographic observations taken during this period shows a dynamic height rise of about 20 cm confined to a zonal band about 100 km wide, that is, the width of the Gulf Stream, and negligible change to the north or south. However, the WOA98v2 analysis shows this 20-cm signal covering a much wider area, including all of the Slope Water region shoreward of the Gulf Stream, between Cape Hatteras and Newfoundland. In contrast, none of the gauges along the east coast of North America show a 20-cm rise in the 1970s. To assess the impact of this error on the calculation by Cabanes *et al.*⁶ of the average warming-induced rise for their nine study areas, assuming a minimal rate of 0.5 mm yr⁻¹ for eight of their regions and 6.7 mm yr⁻¹ for the northeast coast of North America, yields an average rate of 1.2 mm yr⁻¹, which is 2–3 times greater than global estimates of the warming effect⁸. Thus nearly all of the alleged anomalous local effects in the study by Cabanes *et al.*⁶ can be explained by errors in the WOA98v2 analysis.

Our results help to resolve two aspects of the present controversy over twentieth-century GSLR. Concerning the question of whether or not the gauge results are biased high owing to above-average local ocean warming, we conclude that they are not. It follows that the weight of direct evidence strongly suggests a rate of 1.5–2.0 mm yr⁻¹ for twentieth-century GSLR. This is in agreement with most traditional estimates from long gauge records, and is reasonably consistent with the $\sim$ 2.5 mm yr⁻¹ rate obtained by ongoing satellite altimeters for the period 1993–2003.

Concerning the causes of sea level rise, our results provide clear evidence that changes in ocean volume due to temperature and salinity account for only a fraction of sea level change, and that mass change plays a dominant role in twentieth-century GSLR. This aspect of our results is consistent with the results of Antonov *et al.*⁷, who show that the global oceans freshened during the latter half of the twentieth century by an amount equivalent to 1.4 mm yr⁻¹ of fresh water, but goes further by indicating that the source must be continental. The only alternative to this interpretation is that what we identify as mass change is actually a mass redistribution within the global ocean rather than a mass increase due to the addition of fresh water. However, for this to be true, large areas of the global ocean would have to have falling sea levels for the entire twentieth century. Observations from the global tide gauge network, consisting mostly of gauge sites located along the margins of the ocean basins, do

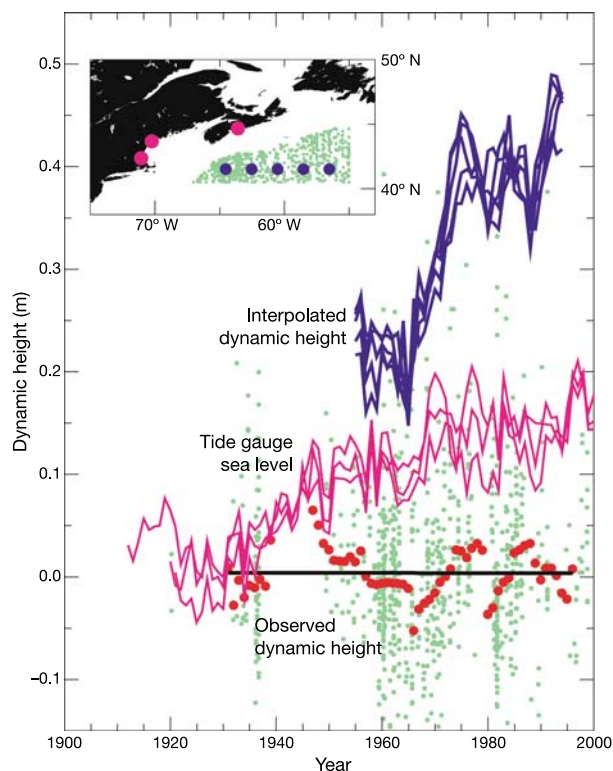


Figure 4 Regional comparison of hydrographic and tide-gauge-measured sea level change in the Western Atlantic. As in Fig. 1, but hydrographic observations limited to 1,300 km $\times$ 500 km triangular region adjacent to tide gauge sites at Boston, Portland and Halifax. Time series of dynamic height anomalies computed from the WOA98v2 interpolated data set are shown in blue, offset by 0.4 m. Map inset shows tide gauge locations in purple, observed dynamic height locations in green and interpolated dynamic height locations in blue.

not support this viewpoint. Whether the mid-oceans are currently undergoing such changes can only be determined from long-term, high-precision satellite altimeter missions, such as the TOPEX/Poseidon and Jason missions, which are at present under way. $\square$

Methods

Data sources

All of the tide gauge records used in this study were obtained from the Permanent Service for Mean Sea Level (PSMSL)¹⁴ in the form of monthly average relative sea level measurements. Each record was corrected for vertical land motion due to glacial isostatic adjustment¹⁵, but not for the inverted barometer effect caused by atmospheric pressure variations. The PSMSL¹⁴ in its table of trends shows formal errors (SD) of about 0.2–0.4 mm yr⁻¹ for records >60 yr. The uncertainty of our trends derived from independent 5-yr gauge means is 0.1–0.3 mm yr⁻¹, less because of the further smoothing beyond 1 yr.

Observed (as opposed to interpolated) hydrographic profile data were obtained from the World Ocean Atlas 1998v2¹² and density (steric) changes computed from both temperature and salinity variability. To avoid over-weighting the data set with shallow expendable bathythermograph (XBT) measurements, only traditional bottle, salinity–temperature–depth (STD), and conductivity–temperature–depth (CTD) measurements to 500 m or more were used. Surface dynamic heights (0 to 500 m, to 1,000 m, and so on) were calculated for each profile by computing specific volume anomalies at observed levels and then integrating in the vertical. In order to correctly average over regions with significant variation of mean dynamic topography, dynamic height deviations were computed relative to a 1° $\times$ 1° (or in data-sparse areas, 2° $\times$ 2°) mean dynamic topography interpolated to the actual location of each observation. These deviations were subsequently converted into dynamic height anomalies by subtracting mean annual and semi-annual cycles based on a least-squares fit to the month-by-month composite of all measurements within a given region.

Weighting and sampling

Because hydrographic observations tend to be concentrated near land or along specific track lines and there is evidence of geographical variations in the steric trends, it is important to consider how area weighting might lead to a different estimate of the trend for a particular region. To this end, we computed 1,000-m dynamic height trends for each 10° $\times$ 10° area in the eastern Pacific study area shown in Fig. 1. The results range between 0.1

and 1.8 mm yr⁻¹ with a regional average value of 0.8 mm yr⁻¹. Although this figure is somewhat higher than the trend computed without area weighting, the difference is not significant considering the poor temporal sampling found in several of the 10° × 10° areas. Thus, we conclude that area weighting is not essential to this study. The formal uncertainties of trends of dynamic height derived from 5-yr means are 0.1–0.2 mm yr⁻¹. These are smaller than the uncertainty values from gauge data because the former are based on area averages.

Except for the Balboa area, which suffers from relatively poor temporal sampling, the hydrographic results presented in Fig. 2 are not overly sensitive to the size or orientation of the sample areas. Cutting the width of the California area in half, that is, setting the offshore boundary at 220 km instead of 440 km as shown, results in a dynamic height trend of 0.4 mm yr⁻¹ instead of 0.5 mm yr⁻¹ over 1938–1996. Similarly, cutting the Honolulu area in half results in a dynamic height trend of 0.4 mm yr⁻¹ instead of 0.3 mm yr⁻¹ from 1950 to 1999.

Received 15 July; accepted 22 December 2003; doi:10.1038/nature02309.

1. Church, J. A. *et al.* *Climate Change 2001: The Scientific Basis* (ed. Houghton, J. T.) Ch. 11 (Cambridge Univ. Press, New York, 2001).
2. Douglas, B. C. & Peltier, W. R. The puzzle of global sea level rise. *Phys. Today* **55**(3), 35–40 (2002).
3. Munk, W. Twentieth century sea level: an enigma. *Proc. Natl Acad. Sci. USA* **99**, 6550–6555 (2002).
4. Meier, M. & Wahr, J. M. Sea level is rising: do we know why? *Proc. Natl Acad. Sci. USA* **99**, 6524–6526 (2002).
5. Munk, W. Ocean freshening, sea level rising. *Science* **300**, 2041–2043 (2003).
6. Cananes, C., Cazenave, A. & LeProvost, C. Sea level rise during past 40 years determined from satellite and in situ observations. *Science* **294**, 840–842 (2001).
7. Antonov, J. I., Levitus, S. & Boyer, T. P. Steric sea level variations 1957–1994: importance of salinity. *J. Geophys. Res.* **107**, doi:10.1029/2000/JC000964 (2002).
8. Levitus, S., Antonov, J. I., Boyer, T. P. & Stephens, C. Warming of the world ocean. *Science* **287**, 2225–2229 (2000).
9. Warrick, R. A., Le Provost, C., Meier, M. F., Oerlemans, J. & Woodworth, P. L. in *Climate Change 1995, The Science of Climate Change* (eds Houghton, J. T. *et al.*) 359–405 (Cambridge Univ. Press, Cambridge, 1996).
10. Nerem, R. S. & Mitchum, G. T. in *Sea Level Rise: History and Consequences* (eds Douglas, B. C., Kearney, M. S. & Leatherman, S. P.) 65–95 (Academic, San Diego, 2001).
11. Ducet, N., Le Traon, P. Y. & Reverdin, G. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.* **C 105**, 19477–19498 (2002).
12. World Ocean Atlas 1998v2. (<http://www.nodc.noaa.gov/OC5/wod98v2.html>)
13. Joyce, T. M., Deser, C. & Spall, M. A. The relation between decadal variability of Subtropical Mode Water and the North Atlantic Oscillation. *J. Clim.* **13**, 2550–2569 (2000).
14. Permanent Service for Mean Sea Level (http://www.pol.ac.uk/psmsl/psmsl_individual_stations.html) (accessed 2003).
15. Peltier, W. R. in *Sea Level Rise: History and Consequences* (eds Douglas, B. C., Kearney, M. S. & Leatherman, S. P.) 65–95 (Academic, San Diego, 2001).

Acknowledgements We thank G. Mitchum for comments and suggestions; T. Rossby, P. Woodworth, S. Gille, R. Cheney, J. Lillibridge, W. Smith and B. Miller for insights; and C. Y. Kuo and C. K. Shum for sharing their computations based on the WOA98 interpolated data set. This work was supported in part by NOAA's Climate Services and Observations Program and NASA. The views, opinions, and findings contained in this report are those of the authors, and should not be construed as an official NOAA or US Government position, policy or decision.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.M. (laury.miller@noaa.gov).

Experimental evidence for the existence of iron-rich metal in the Earth's lower mantle

Daniel J. Frost¹, Christian Liebske¹, Falko Langenhorst¹, Catherine A. McCammon¹, Reidar G. Trønnes^{1,2} & David C. Rubie¹

¹Bayerisches Geoinstitut, University of Bayreuth, D-95440 Bayreuth, Germany

²Nordic Volcanological Institute, Natural Sciences Building, University of Iceland, IS-101, Reykjavik, Iceland

The oxidation state recorded by rocks from the Earth's upper mantle can be calculated from measurements of the distribution of Fe³⁺ and Fe²⁺ between the constituent minerals^{1–3}. The capacity for minerals to incorporate Fe³⁺ may also be a significant factor controlling the oxidation state of the mantle^{4,5}, and high-pressure experimental measurements of this property

might provide important insights into the redox state of the more inaccessible deeper mantle. Here we show experimentally that the Fe³⁺ content of aluminous silicate perovskite, the dominant lower-mantle mineral, is independent of oxygen fugacity. High levels of Fe³⁺ are present in perovskite even when it is in chemical equilibrium with metallic iron. Silicate perovskite in the lower mantle will, therefore, have an Fe³⁺/total Fe ratio of at least 0.6, resulting in a whole-rock ratio of over ten times that of the upper mantle^{5,6}. Consequently, the lower mantle must either be enriched in Fe³⁺ or Fe³⁺ must form by the disproportionation of Fe²⁺ to produce Fe³⁺ plus iron metal. We argue that the lower mantle contains approximately 1 wt% of a metallic iron-rich alloy. The mantle's oxidation state and siderophile element budget have probably been influenced by the presence of this alloy.

The Earth's upper mantle contains both ferric and ferrous iron but ferrous iron is by far the more dominant species (Fe³⁺/ΣFe < 0.03) (refs 5, 6). Chemical equilibria between Fe³⁺- and Fe²⁺-bearing minerals in mantle rocks can be used to calculate the oxygen fugacity of the upper mantle, which is generally found to fall within +1 and –2 log units of the quartz–fayalite–magnetite oxygen buffer^{1–3}. This lower bound is more than three orders of magnitude above the level where the upper mantle would be in chemical equilibrium with metallic iron, as probably occurred during core–mantle segregation.

The oxygen fugacity of an upper-mantle assemblage equates quite reasonably with the concentration of Fe³⁺ in some constituent minerals. The Fe³⁺ content of spinel, for example, decreases with decreasing oxygen fugacity towards zero at conditions of equilibrium with metallic iron³. This relationship is dependent on mineral structure and chemistry, however, and at higher pressures there are examples of silicate minerals (such as wadsleyite, ringwoodite and majoritic garnet) that have significant Fe³⁺ concentrations when coexisting with metallic iron^{4,5}. The crystal–chemical constraints that make Fe²⁺ the dominant iron species in the upper mantle may not, therefore, hold for the transition zone and lower mantle, where mixtures of Fe³⁺-rich minerals and metallic Fe may become stable, even if bulk oxygen contents are the same as in the upper mantle.

Using a multianvil high-pressure apparatus, we have synthesized samples of aluminous silicate perovskite coexisting with metallic Fe to determine the Fe³⁺/ΣFe ratio of perovskite under these relatively reducing conditions. Aluminous pyroxene starting materials were ground with an additional 20 wt% powdered metallic Fe and equilibrated in Fe capsules at 1,600 °C for run durations of up to five days. In most experiments 20 wt% ferropericlase was also added. Additional experiments in graphite capsules using a synthetic peridotite composition, without added metallic Fe, were performed at conditions slightly below and above the silicate solidus. The Fe³⁺/ΣFe ratios of silicate perovskite in the recovered samples were measured using Mössbauer and electron energy loss spectroscopy⁷ and are reported in Table 1. Figure 1a shows a typical run product. In addition to using a Fe capsule, distributing metallic Fe throughout the experimental charge is essential for ensuring redox equilibrium between the silicate and metal.

Figure 2 shows the Fe³⁺ versus the Al³⁺ content of perovskite in atoms per two-cation formula unit for samples equilibrated with metallic Fe at typical mantle temperatures. The shaded region encompasses results with uncertainties from similar experiments performed under more oxidizing conditions in the presence of Re and ReO₂ (refs 8, 9). The Fe³⁺ content of perovskite is coupled to the Al³⁺ concentration^{8–10}. The nonlinear relationship between Al³⁺ and Fe³⁺ may result from a gradual change in the dominant trivalent cation incorporation mechanism—from one involving oxygen vacancies at low Al³⁺ concentrations, to a coupled substitution of trivalent cations at higher Al³⁺ concentrations⁹. Computer simulations indicate that such a coupled substitution remains favourable at least up to 100 GPa, that is, over most of the lower

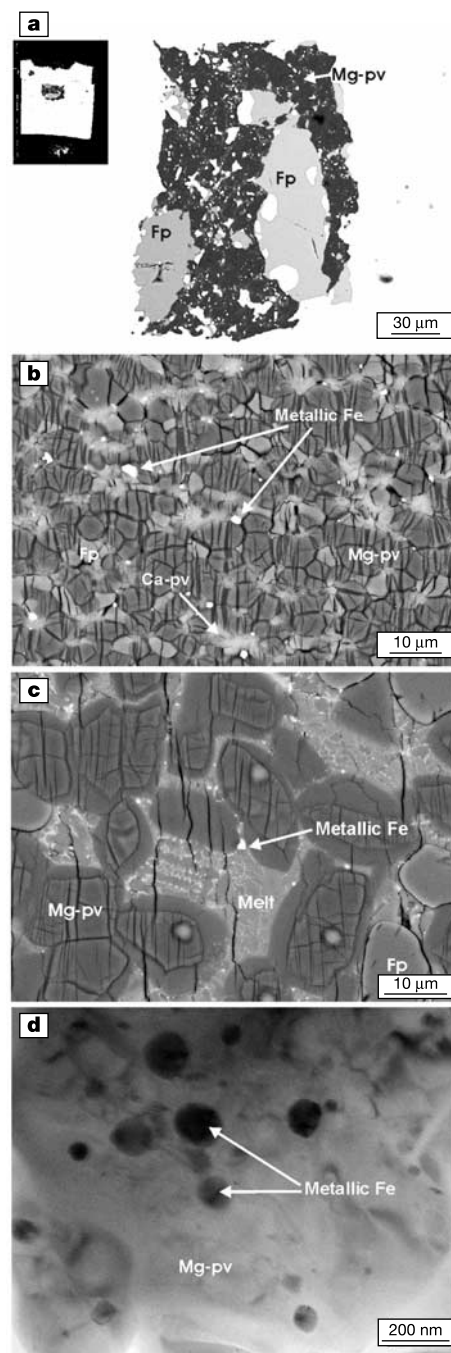


Figure 1 Scanning (a–c) and transmission (d) electron microscope images of experimental run products. **a**, An iron capsule containing magnesium silicate perovskite (Mg-pv) and ferropericlase (Fp). Bright metallic Fe powder, which was added to the starting material, can be seen throughout the sample. Inset, the entire capsule rotated 90°. The capsule is 0.8 mm in diameter. An Al₂O₃ sleeve surrounded the Fe capsule inside the multi-anvil high-pressure assembly. **b**, A sub-solidus run product in a diamond capsule. Grains of magnesium silicate perovskite, calcium perovskite (Ca-pv) and ferropericlase (Fp) can be seen with small amounts of metallic Fe that formed during the experiment. Metallic Fe was not added to these starting materials. **c**, A super-solidus experiment run in a diamond capsule where magnesium silicate perovskite and silicate melt can be seen with a grain of metallic Fe that must also have formed during the experiment. **d**, A TEM image of a cluster of submicrometre-sized particles of metallic Fe inside ferric iron-rich magnesium silicate perovskite (run 2002).

mantle¹¹. Our new data indicate that the relationship between Al³⁺ and Fe³⁺ is independent of oxygen fugacity because the Fe³⁺ content of perovskite in equilibrium with metallic Fe is very similar to the level measured under more-oxidizing conditions. The oxygen fugacity of the Re–ReO₂ buffer is approximately 6 log units above the metallic Fe equilibrium. At oxygen fugacities slightly lower than those investigated here, all Fe in the system would be reduced to metal. Experiments performed above the silicate solidus also demonstrate that perovskite has a significant Fe³⁺ concentration during either crystallization or partial melting (Fig. 2).

If the lower mantle has a typical bulk silicate earth composition^{12,13}, then it will be composed of approximately 70 wt% silicate perovskite, which will contain around 5 wt% Al₂O₃ (refs 14, 15). Using our Fe³⁺ versus Al³⁺ relationship, we calculate that perovskite in the lower mantle must have an Fe³⁺/ΣFe ratio of approximately 0.6. This means that either the lower mantle is enriched in Fe³⁺ compared to the upper mantle, or that silicate perovskite appropriates oxygen either by the reduction of volatile species or by the disproportionation of Fe²⁺ to Fe³⁺ and metallic Fe.

The first possibility is inconsistent with evidence for whole-mantle convection¹⁶. The redox state of the upper mantle seems to have remained relatively constant since 3.6 Gyr ago^{17,18}, which would not have occurred if there were significant mixing with an Fe³⁺-rich lower mantle. The second possibility implies that the bulk oxygen content of the lower mantle is the same or similar to the upper mantle and that oxidation of Fe²⁺ to Fe³⁺ is balanced by reduction of some other component. Volatile species such as H₂O and CO₂ could be reduced to diamond, CH₄ and H₂ in order to release oxygen, but the mantle budget of CO₂ and H₂O together is probably only of the order of 2,000 p.p.m. at most¹⁹, so such a reduction could only account for a fraction of the required oxygen. Several studies have also proposed that in deeper regions of the upper mantle and transition zone these volatile species are already in a reduced state^{5,20,21}. The most realistic scenario, therefore, is for oxygen to be provided by the reduction of Fe²⁺ to metallic iron:

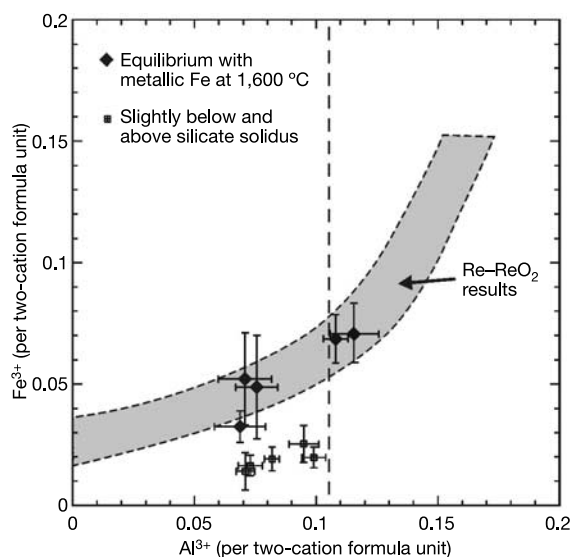


Figure 2 The variation of Al³⁺ and Fe³⁺ in silicate perovskite reported as atoms per two-cation formula unit. Analyses from experiments performed in equilibrium with metallic Fe at a typical temperature for the top of the lower mantle (1,600 °C) are shown as black diamonds. The grey field encompasses analyses from experiments performed at a higher oxygen fugacity in equilibrium with the Re–ReO₂ buffer^{8,9}. Analyses of perovskites forming slightly below and above the silicate solidus (~2,200 °C) are shown as small squares. The vertical dotted line indicates the approximate Al content of silicate perovskite expected in a lower mantle with a typical bulk silicate earth composition^{12,13}.

Table 1 Experimental conditions and perovskite compositions

Run number	Pressure (GPa)	Temperature (°C)	Time (h)	Fe ³⁺ /ΣFe × 100	Si	Al	Fe ³⁺	Fe ²⁺	Mg	Ca
Fe capsules with Fe metal powder added to the starting material										
202	24	1,600	120	48 (5) E	0.927 (10)	0.069 (10)	0.032 (6)	0.035 (7)	0.937 (27)	—
218a	24	1,600	96	35 (15) M	0.922 (11)	0.076 (11)	0.051 (22)	0.094 (22)	0.858 (19)	—
218b	24	1,600	96	37 (16) M	0.923 (69)	0.075 (8)	0.049 (21)	0.083 (21)	0.869 (14)	—
2002	26	1,600	24	68 (10) E	0.897 (16)	0.115 (9)	0.071 (11)	0.034 (8)	0.884 (23)	—
2014	26	1,600	7	66 (10) M	0.898 (11)	0.108 (5)	0.069 (10)	0.035 (10)	0.889 (11)	—
Graphite capsules with no added Fe metal powder										
3224	24	2,150	2	20 (11) M	0.939 (6)	0.071 (4)	0.014 (8)	0.056 (8)	0.887 (9)	0.020 (7)
3242*	24	2,300	0.2	30 (9) M	0.901 (8)	0.095 (6)	0.025 (8)	0.059 (8)	0.884 (11)	0.025 (2)
				46 (10) E	0.91	0.099	0.020	0.023	0.916	0.025
				39 (10) E	0.93	0.073	0.016	0.026	0.927	0.024
				39 (10) M	0.926 (8)	0.082 (3)	0.019 (5)	0.030 (5)	0.911 (11)	0.024 (4)

Letters after Fe³⁺/ΣFe values refer to the analysis technique that is either electron energy loss (E) or Mössbauer (M) spectroscopy. Electron microprobe analyses are normalized to a two-cation formula unit (for example, MgSiO₃). Minor amounts of Ti, Cr, Ni, Mn and Na are excluded from the perovskite compositions of experiments 3224 and 3242.

*As a result of the thermal gradient experiment 3242 encompassed a range of conditions from sub-solidus to the silicate liquidus. Analyses were made at various points along this regime. The first two analyses are from a region close to the solidus, whereas the second two are from regions containing silicate melt.

The observation that high levels of Fe³⁺ can occur in perovskite in equilibrium with metallic Fe means that Fe²⁺ disproportionation must occur if aluminous perovskite is formed in a system that is initially poor in Fe³⁺. The formation of aluminous perovskite with the appropriate Fe³⁺ content from a bulk silicate earth^{12,13} composition with an initial Fe³⁺/ΣFe ratio similar to the upper mantle^{5,6} requires the precipitation of approximately 1 wt% metallic Fe-rich alloy. The alloy would contain approximately 88 wt% Fe, 10 wt% Ni, and 1 wt% S together with the majority of the lower mantle's siderophile elements. The Ni content, calculated using 25 GPa partition coefficients²², may decrease with pressure²³. Although the perovskite Fe³⁺ concentration is slightly lower at the silicate solidus (Fig. 2), approximately 0.5 wt% alloy would still be expected to form at these conditions.

Circumstantial evidence in support of disproportionation can be found in experiments performed on peridotite compositions in diamond capsules. Powdered metallic Fe was not added to these experiments but Fe-rich metal has formed in equilibrium with aluminous silicate perovskite (Fig. 1b, c). Transmission electron microscope images also indicate the formation of submicrometre metal inclusions (Fig. 1d).

This proportion of metal is unlikely to have separated from the solid lower mantle even at temperatures above the metal solidus as a consequence of the high interfacial energies between Fe-rich liquid metal and lower-mantle minerals^{24,25}. During mantle convection, Fe²⁺ in descending material entering the lower mantle would disproportionate to exsolve metal, which would remain locked in the silicate assemblage. As material ascended out of the stability field of perovskite, Fe³⁺ and metal would recombine to Fe²⁺, retaining the same bulk oxygen content. Such a metal phase would probably evade geophysical detection. A metallic phase of this proportion would not significantly influence the electrical conductivity of the lower mantle because it would not be interconnected. In fact, experiments have shown aluminous and ferric iron-rich perovskite to have an electrical conductivity quite consistent with that of the lower mantle²⁶. This amount of metal would also be too small to measurably influence seismic wave velocities in the lower mantle. The Fe³⁺ content may influence the elastic properties of silicate perovskite, however, which should be taken into consideration when comparing lower-mantle seismic velocities with experimentally and theoretically determined values.

The formation of the metallic phase in the lower mantle has important implications for the evolution of the redox state of the mantle. Core-mantle segregation would have ensured that the early silicate mantle was more reduced than present-day upper mantle^{27,28}. Irrespective of whether silicate perovskite formed by solid-state transformation during accretion or by crystallization

from a magma ocean, disproportionation of Fe²⁺ would have occurred concurrently because the precursor material would have been poor in Fe³⁺. Iron-rich metal must, therefore, have existed in the lower mantle from the moment silicate perovskite became stable and could not have separated to a large extent without creating a lower mantle enriched in oxygen.

Many accretion models propose that core-forming liquid metal descended through the solid lower mantle^{23,28,29}. In this case, the descending metal could have transported some of the disproportionated metal into the core. Transport of the metal to the core could also have been aided by the formation of a deep magma ocean from which metal would exsolve as aluminous silicate perovskite crystallized. This would result in a net increase in the oxygen content of the lower mantle and after convection the bulk oxygen content of the mantle would be raised. This is consistent with the observation that the upper mantle today is more oxidized than it was likely to have been during core formation^{3,28}. To increase the Fe³⁺/ΣFe ratio of the upper mantle to its present-day level, from that expected during core formation, requires of the order of 10% of the disproportionated metal phase to segregate into the core.

If a solid lower mantle were in place before the completion of core formation, then the metallic phase could have retained a significant quantity of siderophile elements that would have otherwise been extracted to the core. Subsequent homogenization of the mantle by convection might then help to explain why many siderophile elements are over-abundant in the upper mantle in comparison to their expected behaviour during core-mantle differentiation³⁰. If correct, this eliminates the necessity for a chondritic 'late veneer'²⁷ in geochemical accretion models. □

Received 4 November 2003; accepted 17 February 2004; doi:10.1038/nature02413.

- O'Neill, H. St C. & Wall, V. J. The olivine-orthopyroxene-spinel oxygen geobarometer, the nickel precipitation curve, and the oxygen fugacity of the Earth's upper mantle. *J. Petrol.* **28**, 1169–1191 (1987).
- Luth, R. W., Virgo, D., Boyd, F. R. & Wood, B. J. Ferric iron in mantle derived garnets: implications for thermobarometry and the oxidation state of the mantle. *Contrib. Mineral. Petrol.* **104**, 56–72 (1990).
- Ballhaus, C., Berry, R. F. & Green, D. H. High pressure experimental calibration of the olivine-orthopyroxene-spinel oxygen geobarometer: implications for the oxidation state of the upper mantle. *Contrib. Mineral. Petrol.* **107**, 27–40 (1991).
- O'Neill, H. St C. *et al.* Mössbauer spectroscopy of mantle transition zone phases and determination of minimum Fe³⁺ content. *Am. Mineral.* **78**, 456–460 (1993).
- O'Neill, H. St C. *et al.* Ferric Iron in the Upper Mantle and in Transition Zone Assemblages: Implications for Relative Oxygen Fugacities in the Mantle Vol. 14, 73–89 (Geophys. Monogr. 74, International Union of Geodesy and Geophysics (IUGG), Washington DC, 1993).
- Canil, D. & O'Neill, H. St C. Distribution of ferric iron in some upper-mantle assemblages. *J. Petrol.* **37**, 609–635 (1996).
- van Aken, P. A., Liebscher, B. & Styrsky, V. S. Quantitative determination of iron oxidation states in minerals using Fe L_{2,3}-edge electron energy-loss near-edge structure spectroscopy. *Phys. Chem. Mineral.* **5**, 323–327 (1998).
- Lauterbach, S., McCammon, C. A., van Aken, P., Langenhorst, F. & Seifert, F. Mössbauer and ELNES spectroscopy of (Mg,Fe)(Si,Al)O₃ perovskite: a highly oxidised component of the lower mantle. *Contrib. Mineral. Petrol.* **138**, 17–26 (2000).

9. Frost, D. J. & Langenhorst, F. The effect of Al_2O_3 on Fe-Mg partitioning between magnesiowüstite and magnesium silicate perovskite. *Earth Planet. Sci. Lett.* **199**, 227–241 (2002).
10. McCammon, C. Perovskite as a possible sink for ferric iron in the lower mantle. *Nature* **387**, 694–696 (1997).
11. Richmond, N. C. & Brodholt, J. P. Calculated role of alumina in the incorporation of ferric iron into magnesium silicate perovskite. *Am. Mineral.* **83**, 947–951 (1998).
12. McDonough, W. F. & Sun, S.-s. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
13. O'Neill, H. S. C. & Palme, H. in *The Earth's Mantle* (ed. Jackson, I.) 3–126 (Cambridge Univ. Press, Cambridge, 1998).
14. Irifune, T. Absence of an aluminous phase in the upper part of the lower mantle. *Nature* **370**, 131–133 (1994).
15. Kesson, S. E., Fitz Gerald, J. D. & Shelley, J. M. Mineralogy and dynamics of a pyrolite lower mantle. *Nature* **392**, 252–255 (1998).
16. van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
17. Canil, D. Vanadium partitioning and the oxidation state of Archaean komatiite magmas. *Nature* **389**, 842–845 (1997).
18. Delano, J. W. Redox history of the Earth's interior since ~3900 Ma: Implications for prebiotic molecules. *Orig. Life Evol. Biosph.* **31**, 311–341 (2001).
19. Wood, B. J., Pawley, A. & Frost, D. J. Water and carbon in the Earth's mantle. *Phil. Trans. R. Soc. Lond. A* **354**, 1495–1511 (1996).
20. Ballhaus, C. Is the upper-mantle metal-saturated? *Earth Planet. Sci. Lett.* **132**, 75–86 (1995).
21. Ballhaus, C. & Frost, B. R. The generation of oxidized CO_2 -bearing basaltic melts from reducing CH_4 -bearing upper-mantle sources. *Geochim. Cosmochim. Acta* **58**, 4931–4940 (1994).
22. Asahara, Y., Kubo, T. & Kondo, T. Phase relations of a carbonaceous chondrite at lower mantle conditions. *Phys. Earth Planet. Inter.* (in the press).
23. Walter, M. J., Newsom, H. E., Ertel, W. & Holzheid, A. in *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 285–289 (Univ. Arizona Press, Tucson, 2000).
24. Shannon, M. C. & Agee, C. B. Percolation of core melts at lower mantle conditions. *Science* **280**, 1059–1061 (1998).
25. Agee, C. B. & Shannon, M. C. Experimental constraints on percolative core formation at lower mantle conditions. *Lunar Planet. Sci. Abstr.* **XXVIII**, 7 (1997).
26. Xu, Y. S., McCammon, C. & Poe, B. T. The effect of alumina on the electrical conductivity of silicate perovskite. *Science* **282**, 922–924 (1998).
27. O'Neill, H. S. C. The origin of the Moon and the early history of the Earth—A chemical model. Part 2: The Earth. *Geochim. Cosmochim. Acta* **55**, 1159–1172 (1991).
28. Righter, K., Drake, M. J. & Yaxley, G. Prediction of siderophile element metal-silicate partition coefficients to 20 GPa and 2,800 °C: the effect of pressure, temperature, f_{O_2} and silicate and metallic melt composition. *Phys. Earth Planet. Inter.* **100**, 115–134 (1997).
29. Stevenson, D. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 231–249 (Oxford Univ. Press, Oxford, 1990).
30. Ringwood, A. E. Chemical evolution of the terrestrial planets. *Geochim. Cosmochim. Acta* **30**, 41–104 (1966).

Acknowledgements We thank H. Fischer, G. Herrmannsdörfer, D. Krause and H. Schulze for technical assistance. The German Science Foundation (DFG) and the EU Access to Research Infrastructures Programme supported this research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.F. (dan.frost@uni-bayreuth.de).

Antibiotic-mediated antagonism leads to a bacterial game of rock–paper–scissors *in vivo*

Benjamin C. Kirkup & Margaret A. Riley

Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06511, USA

Colicins are narrow-spectrum antibiotics produced by and active against *Escherichia coli* and its close relatives. Colicin-producing strains cannot coexist with sensitive or resistant strains in a well-mixed culture, yet all three phenotypes are recovered in natural populations¹. Recent *in vitro* results conclude that strain diversity can be promoted by colicin production in a spatially structured, non-transitive interaction², as in the classic non-transitive model rock–paper–scissors (RPS). In the colicin version of the RPS model, strains that produce colicins (C) kill

sensitive (S) strains, which outcompete resistant (R) strains, which outcompete C strains. Pairwise *in vitro* competitions between these three strains are resolved in a predictable order (C beats S, S beats R, and R beats C), but the complete system of three strains presents the opportunity for dynamic equilibrium². Here we provide conclusive evidence of an *in vivo* antagonistic role for colicins and show that colicins (and potentially other bacteriocins) may promote, rather than eliminate, microbial diversity in the environment.

Colicins are high-molecular-mass bacteriocins produced by *E. coli*^{3,4}. *In vitro* experiments show that coexistence of a colicin producer and a sensitive strain in a well-mixed culture is not possible^{5–7}. In natural populations a further phenotype, colicin-resistant, is always present, and more recent *in vitro* and *in silico* work has incorporated this third phenotype^{2,6,8,9}. Theoretical and empirical studies reveal that a combination of colicinogenic, sensitive and resistant provides a stabilizing factor in a mixed, spatially structured assemblage^{2,6,8}. This complex of three strains has the same formal structure as the RPS game², with C beating S (by killing), S beating R (by growth-rate advantage, because the resistant strain has had to modify a receptor used in nutrient uptake), and R beating C (because colicin production involves bacterial suicide). Thus, these authors conclude that antagonistic factors, such as colicins, may promote biological diversity in natural microbial populations and communities².

Colicins have repeatedly been the subject of *in vivo* studies that failed to confirm expectations generated by laboratory and mathematical studies for potent antagonism^{4,10}. However, whereas those studies generally regarded resistance as a nuisance phenomenon, the RPS model treats it as central to the establishment of a dynamic equilibrium. To validate this new model of colicin-related population dynamics, a new *in vivo* experimental model was applied. Controlled enteric bacterial populations were established in small populations of mice so that the bacterial population dynamics could be monitored as bacteria migrated from mouse to mouse.

This mouse model has proved to be highly effective in studies of enteric bacteria in the mammalian colon¹¹. Streptomycin eliminates from the mouse colon the minority Gram-negative microbial community, which is replaced with the desired streptomycin-resistant Gram-negative test strains. The majority Gram-positive communities remain relatively undisturbed^{12,13}, and the mouse experiences normal colon development and function^{11,14}.

To study the impact of colicins on bacterial dynamics in this mouse model, four related *E. coli* strains were generated. The base strain, *E. coli* K12 (BZB1011), is naturally sensitive to most known colicins. A streptomycin-resistant mutant of BZB1011 was selected as the colicin-sensitive (S) strain. Two colicin-producing strains (C_{E1} and C_{E2}) were derived by introducing naturally occurring colicin plasmids into the S strain by electroporation. These two colicins (E1 and E2) were chosen because they represent colicins commonly found in mouse colon isolates, are encoded on easily manipulated, non-transformable, low-molecular-mass plasmids, and act in distinctly different ways. Colicin E2 is a DNase that enters the cell and nonspecifically cleaves DNA¹⁵. In contrast, colicin E1 produces voltage-gated pores in the cell membrane, depolarizing the cell¹⁵. To produce the fourth strain, S was exposed to colicin E2, and a resistant strain (R) was isolated. This strain was resistant to both colicin E1 and colicin E2. DNA sequencing of the *btuB* locus of the R strain suggests that the resistance phenotype was generated by an IS1 insertion (J. E. Wertz and C. Winkworth, personal communication).

Seventy-two mice were given streptomycin in their drinking water. The sensitive and resistant *E. coli* were inoculated by mouth into 24 mice each, and C_{E1} and C_{E2} were inoculated into 12 mice each. In all cases the new strains rapidly colonized their hosts, and all achieved final densities of about 10^6 colony-forming units per gram of faeces, similar to the frequency of enteric bacteria

generally present in the mouse colon¹⁶. These strains were stably maintained for over four weeks.

Experiments with each of the colicin-producing strains were conducted according to a parallel design. For each experiment, 12 cages of three mice each (36 mice per experiment, described above) were established. Each cage served as a model community, and each of the three mice in it initially carried one of three strains (C, S or R). The mice were permitted to interact freely with each other, and the bacterial populations recovered from their faecal pellets were monitored.

In the colicin E1 experiment, changes in the dominant bacterial populations occurred repeatedly in the majority of the mice (Fig. 1). Sampling over 12 weeks at half-weekly intervals showed over 98 putative transitions between the dominant strains. Only rarely was a mixed population recovered from individual faecal pellets; at any given time each mouse was dominated by a single strain. Detection limits for a second strain were 1% of the population. Sampling over 12 weeks (Fig. 1) revealed that strain C_{E1} tends to displace S, R displaces C_{E1}, and S displaces R, as predicted from earlier *in vitro* studies² and the RPS model^{17,18}. Within certain cages S persisted, whereas other cages fixed on the R phenotype. No cages fixed on the C phenotype, which in all but two cages was eliminated by week 7.

An explicitly cooperative RPS model (for example, two R mice yielding a fourfold advantage to R invading the third mouse in that cage) can be tested by examining pairs of opposing hypotheses (for example, S replaces R versus the hypothesis that R replaces S). The RPS cycle comprises three hypotheses, and their individual inverses thus form the null hypothesis that such a cycle does not exist. A conservatively administered one-tailed Wald test allowed rejection (at the 95% confidence level) of all three null hypotheses: that C_{E1} did not displace S, R did not displace C_{E1}, and S did not displace R. The data thus support an RPS model of strain displacement *in vivo*.

The colicin E2 experiment also revealed cycling between the three phenotypes (S, C and R). Within a week, four of 36 mice had already experienced displacement of their dominant strain (Fig. 2). Overall, fewer strain transitions were observed in the colicin E2 experiment (43 versus 123), and the S strain, once eliminated, did not reappear, as it often did in the colicin E1 experiment. Using the above model, the hypothesis that C_{E2} did not displace S was rejected at 98.7%, and the hypothesis that R did not displace C_{E2} was rejected at 97.7%. However, in the colicin E2 experiment, there was insufficient evidence to accept or reject the hypothesis that R replaced S.

In the colicin E2 experiment, the relatively short resident time of S in any of the cages appears to preclude adequate investigation of the R-to-S transition. To investigate this, after all of the mice in the

colicin E2 experiment had been dominated by an R strain, a single new S-equilibrated mouse was swapped into each cage in exchange for one of the three R-equilibrated mice. In the absence of C_{E2}, S invaded and displaced the R strains in 11 of 16 mice (Fig. 3). In this case, a one-tailed Fisher's exact test suggests that S displaces R, with a *P* value of 0.055 (marginally non-significant). Alternatively, if invasions are non-cooperative, on a per cage basis one might assume fixation at S to occur one-third of the time, yet five out of seven fix on R. This is almost significant, and if invasions are presumed to be cooperative, then S becomes a highly significant invader.

The generation of an RPS game in normal form from the data sets is desirable but difficult. Generating models with 6–12 parameters has been pursued, and heuristic maximum-likelihood estimation does find local minima conforming to the expectations of a generalized RPS model. However, at one extreme overfitting is a problem, and at the other the assumptions become questionable. Estimating the six key parameters of an RPS game from the data requires assumptions about cooperativity and the reappearance of strains that had disappeared from a cage. Cooperativity can be included as three additional parameters in a model. However, the reappearance of strains is more difficult to address explicitly. It would require at least three parameters, and the form of these parameters depends on the presumed mechanism by which the strains reappear. One assumption is that these strains are re-acquired from environmental refugia (faeces, bedding). Another is that they fall below the detection threshold while still in the mouse itself, or thrive in a part of the gastrointestinal tract that does not encourage their detection in faeces. A third hypothesis is that they arise by mutation. This is most likely in the cases of resistance (which reappears nine times), and possible for sensitivity (nine times), but highly unlikely for colicin production (which reappears 14 times and whose detection was the least subject to error). These data and the apparent stability of the individual strains in the mice suggest that mutation is not the regenerative method. Furthermore, the low survivorship of strains in dry faeces (data not shown) suggests that undetected minority populations in the mice may stage unexpected reprisals.

These results provide the first conclusive demonstration that colicins are effective antagonistic agents within *E. coli* populations in an animal host. Previous *in vivo* experiments failed to detect any strain transition effect that could be attributed solely to the action of colicin-producing strains^{4,19}. The present set of experiments suggests that colicins E1 and E2 have a significant role in mediating strain dynamics in the mouse colon. The data also suggest the existence of an RPS relationship between C, S and R strains of *E. coli*—as predicted *in vitro* and *in silico*^{2,6,20}. In this study, the pairwise relationships predicted from the RPS model were clearly observed in the mouse colons; that is, on average, C strains displaced S, R displaced C, and S displaced R. However, overall strain diversity

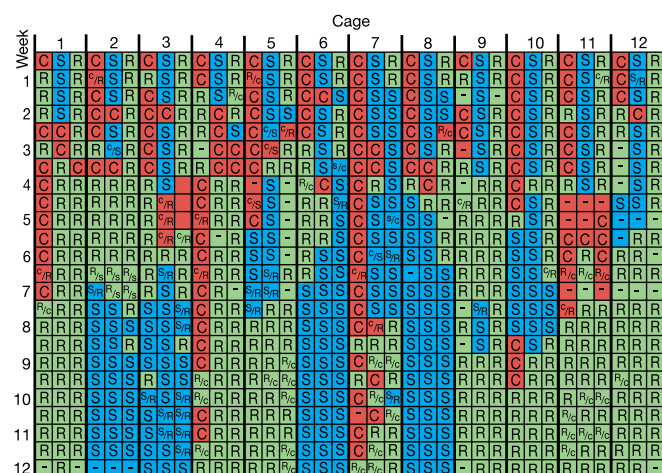


Figure 1 Occupation of co-caged mice by dominant strain: colicin E1. The boxes are coloured to show the dominant strain occupying each mouse. Red represents colicinogenic (C_{E1}); green, resistant; blue, sensitive.

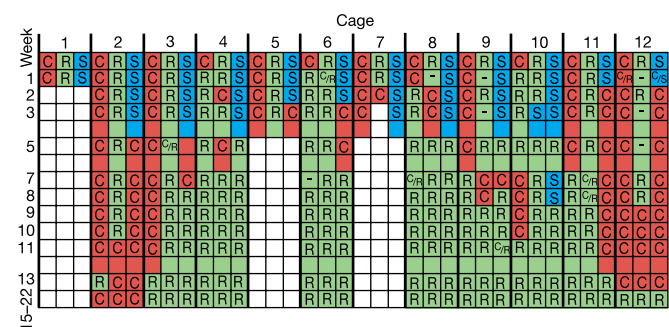


Figure 2 Occupation of co-caged mice by dominant strain: colicin E2. The boxes are coloured to show the dominant strain occupying each mouse. Red represents colicinogenic (C_{E2}); green, resistant; blue, sensitive. Cages 1, 5 and 7 were discontinued because of mice fighting.

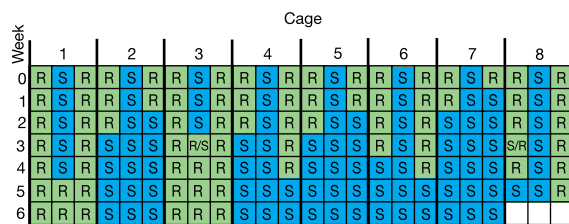


Figure 3 Occupation of co-caged mice by dominant strain after S is reintroduced. The boxes are coloured to show the dominant strain occupying each mouse. Green represents resistant; blue, sensitive. Cage 8 was terminated during week 6 because of a mouse death (natural causes).

was rapidly lost from individual mice, more slowly lost from a cage of mice, and even more slowly lost from all cages.

According to the *in silico* model of colicin-mediated dynamics explored by Kerr *et al.*², strain diversity maintained in the colicin-based RPS system relies not only on the existence of an equilibrium point in the game between the three strains, but also on the existence of numerous local populations. This reduces the probability that a local surge in one of the strains will eliminate any competitor entirely. Kerr *et al.*² were attempting to explain the difference in results obtained from batch culture experiments, in which strain diversity was rapidly lost (S within ~10 generations/one day, C within ~50 generations/five days), and results obtained from agar plate experiments, in which strain diversity was maintained (>70 generations/seven days). They proposed that the local interactions allowed by the structure introduced on the plates stabilized the strain diversity, and they demonstrated that a partial disruption of the structure also destabilized the diversity.

Kerr *et al.* used a model system much more favourable to diversity than the mouse system. The *E. coli* populations in the mouse colon are relatively small (~10⁷ cells per mouse), and the timescales studied in this mouse experiment were an order of magnitude longer (>10 weeks). Kerr *et al.* modelled at least 32 local populations on each plate, and hundreds in each *in silico* world. In contrast, the mice were not populated for extended periods with mixed populations. Thus, each mouse was an 'effective neighbourhood', and each cage contained only three neighbourhoods. As a system becomes discrete, an RPS system maintains diversity under more narrow parameters and for a shorter time¹⁸. Thus, both the length of the experiment and the small size of the populations predisposed the *in vivo* experimental system to resolution rather than diversity. However, in natural mouse populations, such local extinction is not predicted or observed¹. Mice in the wild are considered to have high population structure, in territorial demes of 5–100 individuals^{21,22}, with relatively little interdemographic migration.

The mouse model used here has allowed a significant advance in our understanding of the natural role of microbial toxins in mediating strain and, potentially, species dynamics in natural populations and communities of microbes. We present here the first experimental evidence for an *in vivo* role for colicins, which suggests that this abundant and diverse family of antibiotics serves to promote microbial diversity in one of the dominant niches exploited by enteric bacteria, the mammalian colon. The model does suffer from one significant limitation: the number of mice that can be experimentally monitored precludes adequate modelling of the normal population structure of the mice. This limitation may have precluded the long-term maintenance of strain diversity that has been shown clearly *in vitro*, *in silico* and in natural populations of mice. With the advent of fluorescently labelled strains, it may now be possible to move the experimental venue from the laboratory and animal facilities to a more natural setting by following strain dynamics in wild-caught mouse populations. □

Methods

Strain construction

Strain BZB1011 (ref. 23) was selected for resistance to 100 µg ml⁻¹ streptomycin sulphate (Sigma) to become strain S. Strain C_{E1} was derived from S by electroporation of colicin E1 plasmid, pColE1-K53 (ref. 23), according to standard methods²⁴. Strain C_{E2} was derived from S by electroporation of colicin E2 plasmid, pColE2-p9 (ref. 23), according to standard methods²⁴. Colicin E1 and Colicin E2 lysates were prepared following standard procedures²⁵. Strain R was derived from S by selection for resistance to colicin E2 lysate (800 µl per plate during the initial selection).

Mouse model

Adult CD-1 male mice were obtained from Charles River Laboratories. They were given 5 g per litre streptomycin sulphate for two days before the first inoculation¹⁰. The mice were screened for faecal enteric bacteria by plating faecal pellets on MacConkey agar (Sigma) with streptomycin sulphate. Bacterial inoculations of 10⁹ cells per ml in phosphate-buffered saline (PBS; Sigma) were given by mouth. Faecal samples were taken in sterile plastic boxes, transferred into PBS, massed, homogenized, and then diluted for plating onto MacConkey plates with 100 µg ml⁻¹ streptomycin. Strain phenotypes were determined by velvet transfer of colonies onto LB plates containing 100 µg ml⁻¹ streptomycin and 100 µl colicin lysate, which allows only R and C_{E1} or C_{E2} cells to grow, or onto LB plates containing a strain S cell lawn, which reveals C strains by their killing phenotype.

Wald test

The one-tailed Wald test²⁶ was applied to transition matrices tabulated from the raw data. Some information was lost because the matrices were tabulated as a 10 × 10 matrix, with individual mouse identity homogenized between sample points. The Wald test was conducted with a presumed exponential cooperativity among mice sharing a single cage.

Received 5 December 2003; accepted 17 February 2004; doi:10.1038/nature02429.

- Gordon, D. M., Riley, M. A. & Pinou, T. Temporal changes in the frequency of colicinogeny in *Escherichia coli* from house mice. *Microbiology* **144**, 2233–2240 (1998).
- Kerr, B., Riley, M., Feldman, M. & Bohannan, B. Local dispersal and interaction promote coexistence in a real life game of rock–paper–scissors. *Nature* **418**, 171–174 (2002).
- Fredericq, P. Colicins. *Annu. Rev. Microbiol.* **11**, 7–22 (1957).
- Hardy, K. G. Colicinogeny and related phenomena. *Bacteriol. Rev.* **39**, 464–515 (1975).
- Chao, L. & Levin, B. R. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl Acad. Sci. USA* **78**, 6324–6328 (1981).
- Durrett, R. & Levin, S. Allelopathy in spatially distributed populations. *J. Theor. Biol.* **185**, 165–171 (1997).
- Gordon, D. M. & Riley, M. A. A theoretical and empirical investigation of the invasion dynamics of colicinogeny. *Microbiology* **145**, 655–661 (1999).
- Czárán, T. L., Hoekstra, R. & Pagie, L. Chemical warfare between microbes promotes biodiversity. *Proc. Natl Acad. Sci. USA* **99**, 786–790 (2002).
- Durrett, R. & Levin, S. Spatial aspects of interspecific competition. *Theor. Popul. Biol.* **53**, 30–43 (1998).
- Wadolowski, E. A., Laux, D. C. & Cohen, P. S. Colonization of the streptomycin-treated mouse large intestine by a human fecal *Escherichia coli* strain: role of growth in mucus. *Infect. Immun.* **56**, 1030–1035 (1988).
- Hentges, D. J., Pongpech, P. & Que, J. How streptomycin treatment compromises colonization resistance against enteric pathogens in mice. *Microb. Ecol. Health Dis.* **3**, 105–112 (1990).
- Myhal, M., Laux, D. & Cohen, P. Relative colonizing abilities of human fecal and K-12 strain of *Escherichia coli* in the large intestines of streptomycin-treated mice. *Eur. J. Clin. Microbiol.* **1**, 186–192 (1982).
- Myhal, M., Cohen, P. & Laux, D. Altered colonizing ability for mouse large intestine of a surface mutant of a human faecal isolate of *Escherichia coli*. *J. Gen. Microbiol.* **129**, 1549–1558 (1983).
- Que, J. U., Casey, S. W. & Hentges, D. J. Factors responsible for increased susceptibility of mice to intestinal colonization after treatment with streptomycin. *Infect. Immun.* **53**, 116–123 (1986).
- Lau, C. Y. *Y. Biology* 89 (Yale Univ., New Haven, 1976).
- Wilson, K. H. in *Gastrointestinal Microbes and Host Interactions* (eds Roderick, I., Mackie, B. A. W. & Isaacson, R. E.) 680 (Chapman and Hall, New York, 1997).
- Smith, M. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
- Weibull, J. W. *Evolutionary Game Theory* (The MIT Press, Cambridge, Massachusetts, 1995).
- Sweeney, N.-J. *et al.* The *Escherichia coli* K-12 *gntP* gene allows *E. coli* F-18 to occupy a distinct nutritional niche in the streptomycin-treated mouse large intestine. *Infect. Immun.* **64**, 3497–3503 (1996).
- Feldgarden, M. & Riley, M. A. The phenotypic and fitness effects of colicin resistance in *Escherichia coli* K-12. *Evolution* **53**, 1019–1027 (1999).
- Berry, R. J. Population dynamics of the house mouse. *Symp. Zool. Soc. Lond.* **47**, 395–425 (1981).
- Rowe, F. P. Wild house mouse biology and control. *Symp. Zool. Soc. Lond.* **47**, 575–589 (1981).
- Pugsley, A. P. *Escherichia coli* K-12 strains for use in the identification and characterization of colicins. *J. Gen. Microbiol.* **131**, 369–376 (1985).
- Dower, W., Miller, J. & Rasdale, C. High efficiency transformation of *Escherichia coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6146 (1988).
- Pugsley, A. P. & Oudega, B. in *Plasmids, a Practical Approach* (ed. Hardy, K. G.) 105–161 (IRL Press, Oxford, 1987).
- Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* **54**, 426–482 (1943).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.C.K. (bckirkup@post.harvard.edu).

Myosin gene mutation correlates with anatomical changes in the human lineage

Hansell H. Stedman^{1,3}, Benjamin W. Kozyak¹, Anthony Nelson¹, Danielle M. Thesier², Leonard T. Su¹, David W. Low^{1,5}, Charles R. Bridges¹, Joseph B. Shrager^{1,3}, Nancy Minugh-Purvis^{2,4,5} & Marilyn A. Mitchell¹

¹Department of Surgery and ²Cell and Developmental Biology, ³the Pennsylvania Muscle Institute, School of Medicine, and ⁴Department of Anatomy and Cell Biology, School of Dental Medicine, University of Pennsylvania, and ⁵Division of Plastic Surgery, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA

Powerful masticatory muscles are found in most primates, including chimpanzees and gorillas, and were part of a prominent adaptation of *Australopithecus* and *Paranthropus*, extinct genera of the family Hominidae^{1,2}. In contrast, masticatory muscles are considerably smaller in both modern and fossil members of *Homo*. The evolving hominid masticatory apparatus—traceable to a Late Miocene, chimpanzee-like morphology³—shifted towards a pattern of gracilization nearly simultaneously with accelerated encephalization in early

*Homo*⁴. Here, we show that the gene encoding the predominant myosin heavy chain (*MYH*) expressed in these muscles was inactivated by a frameshifting mutation after the lineages leading to humans and chimpanzees diverged. Loss of this protein isoform is associated with marked size reductions in individual muscle fibres and entire masticatory muscles. Using the coding sequence for the myosin rod domains as a molecular clock, we estimate that this mutation appeared approximately 2.4 million years ago, predating the appearance of modern human body size⁵ and emigration of *Homo* from Africa⁶. This represents the first proteomic distinction between humans and chimpanzees that can be correlated with a traceable anatomic imprint in the fossil record.

We obtained a DNA sequence by degenerate polymerase chain reaction (PCR) that suggested the existence of a hitherto unrecognized human sarcomeric myosin gene (*MYH16*, see Supplementary Information). An unannotated accession file posted during the finalization of the human genome sequence (AC112711) contained the only exact match to the query sequence on a BLAST search. Here, we use homology to other sarcomeric myosins to annotate this locus and generate the complete sequence of a hypothetical human sarcomeric myosin distantly related to the other known members of this subfamily (Fig. 1; see also Supplementary Information). This reconstruction suggested that a frameshift deletion at codon 660 of the messenger RNA of the deduced polypeptide was

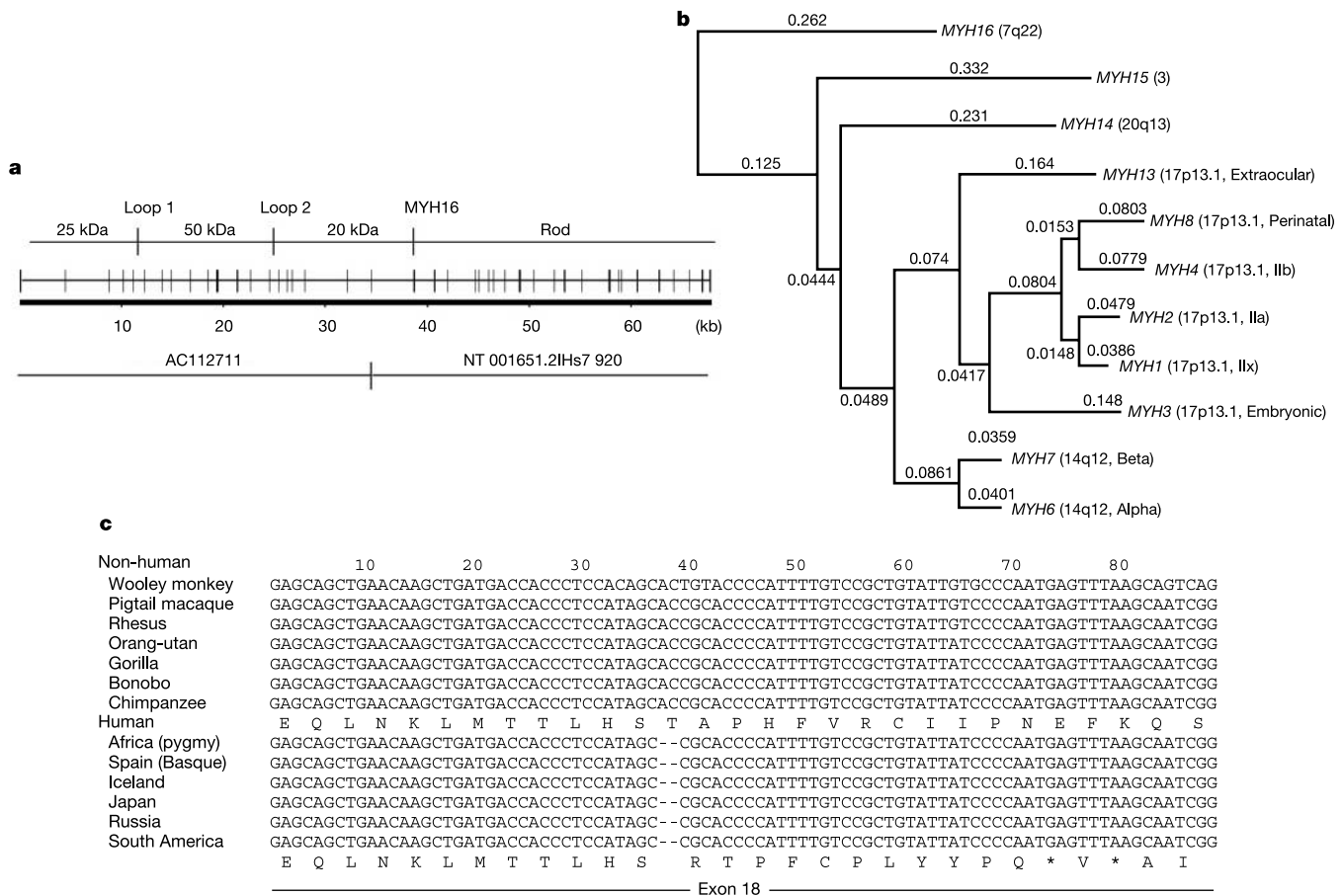


Figure 1 Molecular evolution of *MYH16*. **a**, Distribution of 42 predicted coding exons spanning 67,983 base pairs (bp) in the region of human chromosome 7q22 flanked 5' by *SMURF1* and 3' by *ARPC1A*. **b**, Phylogenetic reconstruction for all human sarcomeric myosin genes (heavy chain), showing early divergence of *MYH16* from others. Branch lengths shown are derived from a maximum likelihood analysis of the aligned cDNAs, beginning with the conserved proline at the head-rod junction. Non-sarcomeric class II

myosins (designated *MYH9*, -10 and -11; data not shown) are used to root the tree. **c**, Aligned DNA sequences for *MYH16* exon 18 representing seven non-human primate species and six geographically dispersed human populations, revealing the effect of frameshift on reading frame and deduced amino acid sequence. Note stop codon at position 72–74.

either a rare allele or a sequencing artefact. To resolve this uncertainty, we sequenced this region in DNA samples from members of geographically disparate human populations. As shown in Fig. 1c, the mutation is found in all modern humans sampled, including natives of Africa, South America, Western Europe, Iceland, Japan and Russia; thus, the inactivating mutation seems to be fixed in *Homo sapiens*. In contrast, all of the non-human primates for which sequence was obtained have an ACC codon that encodes a highly conserved threonine. The frameshift in the human coding sequence truncates the predicted 224-kDa myosin heavy chain to a 76-kDa fragment containing an unstable portion of the myosin head domain (Fig. 1c).

Animals homozygous for null mutations in sarcomeric myosin heavy chain genes sometimes possess severe functional defects in individual muscles that ordinarily accumulate the highest levels of specific isomyosins^{7,8}. In reconstructing the pattern of expression of the *MYH16* gene in a recent common ancestor, scarcity of chimpanzee tissue led us to examine transcription in a wide range of muscles from macaque (*Macaca fascicularis*) and modern humans. Cross-species sequence similarity facilitated the design and use of several isoform-specific RT-PCR primers, with the finding that the amplification products uniquely contain the sequences predicted by our human *MYH16* gene annotation (Fig. 2a). Transcription is detectable only in the muscles of the head, specifically those derived from the embryonic first pharyngeal arch, including temporalis and tensor veli palatini (Fig. 2b).

In gross anatomical comparisons between humans versus great apes and monkeys, the relative size of individual masticatory muscle homologues contrasts remarkably, as is evident in Fig. 3a–k. The relative sizes of the temporalis and masseter muscles are reflected in the morphology of such craniofacial features as the temporal fossa and zygomatic arch (highlighted). Figure 3a–i illustrates the difference in these sites of bony attachment in modern macaque, gorilla and human skulls. At the histological level, the difference between

human and non-human primate temporalis muscle is highlighted by staining for type II (all fast twitch) sarcomeric myosin and interstitial laminin (Fig. 3j–l). Although the unstained type I/slow fibres are nearly identical in cross-sectional area, as confirmed in a reciprocal stain (data not shown), the type II fibres of *H. sapiens* are about one-eighth the size of those of *M. fascicularis*. The relative hypotrophy of the type II fibres resembles that seen in limb muscle in inclusion body myopathy-3 (IBM3)—a rare disease for which the only proven molecular aetiology is a myosin gene mutation (*MYH2*)⁹. Protein gel electrophoresis shows that the dominant myosin heavy chain isoform in the temporalis of *M. fascicularis* is specific to the masticatory muscles and is undetectable in the human temporalis (Fig. 3m). Peptide sequencing of this protein using mass spectrometry unambiguously establishes it as the product of the *M. fascicularis* *MYH16* orthologue (Supplementary Information), whereas the major isomyosins of the human temporalis have recently been identified as products of the *MYH1*, -2 and -7 genes¹⁰.

The fact that the human *MYH16* locus is still appropriately transcribed indicates that the coding sequence deletion identified here was not preceded by a silencing mutation in a transcriptional control domain. We postulate that the volume of the skeletal muscle fibres expressing the *MYH16* gene transcript is proportional to the total amount of myosin heavy chain accumulating in the cell, and that reliance on translation of the less abundant *MYH1* and -2 transcripts in the face of a frameshift mutation in *MYH16* has resulted in the eightfold reduction in the size of the type II fibres in the human masticatory muscles as compared with macaque. Genetic manipulation of muscle size has marked secondary effects on the anatomy of bony attachment sites, as illustrated by recent studies of myostatin signalling^{11,12}. Moreover, experimental animal models of masticatory muscle resection or transposition have demonstrated the correlation between craniofacial morphology and the force of masticatory muscle contraction^{13,14}. We reason

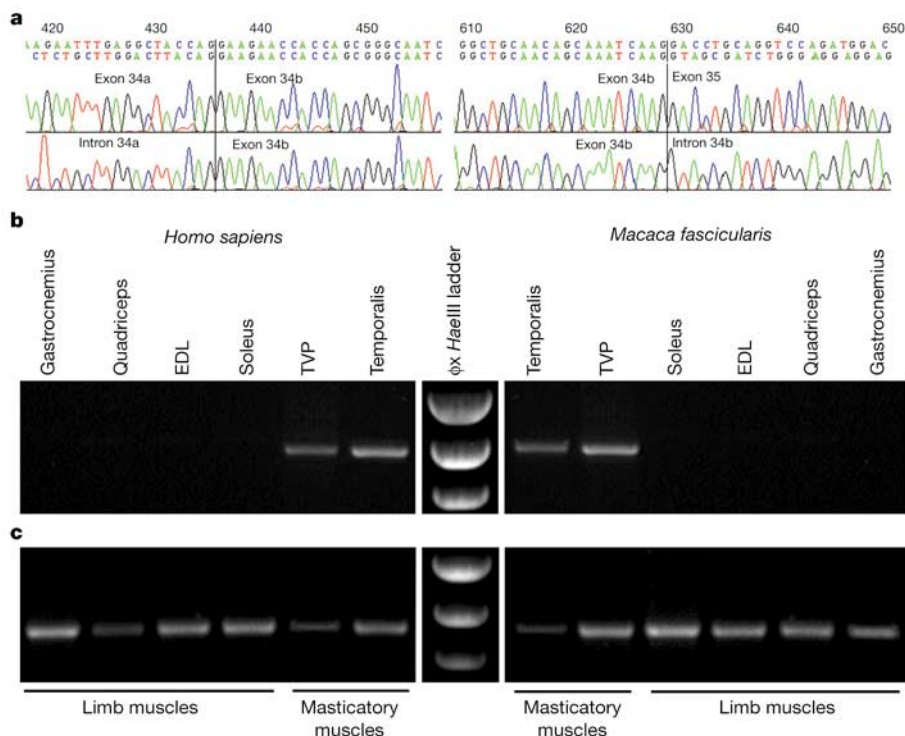


Figure 2 Transcription of *MYH16*. **a**, Alignment of representative genomic and cDNA sequences demonstrate that *MYH16* is a transcribed and properly spliced pseudogene in *H. sapiens*. **b**, RT-PCR detects transcription of *MYH16* and its *M. fascicularis* orthologue only in muscles derived from the first pharyngeal arch (denoted as masticatory muscles).

c, Control amplifications were performed using identical first-strand cDNA template and a pair of 'universal' sarcomeric myosin heavy chain primers based on an MYH consensus sequence. EDL, extensor digitorum longus; TVP, tensor veli palatini; ϕ X HaeIII ladder, ϕ X genomic DNA digested with the restriction enzyme HaeIII.

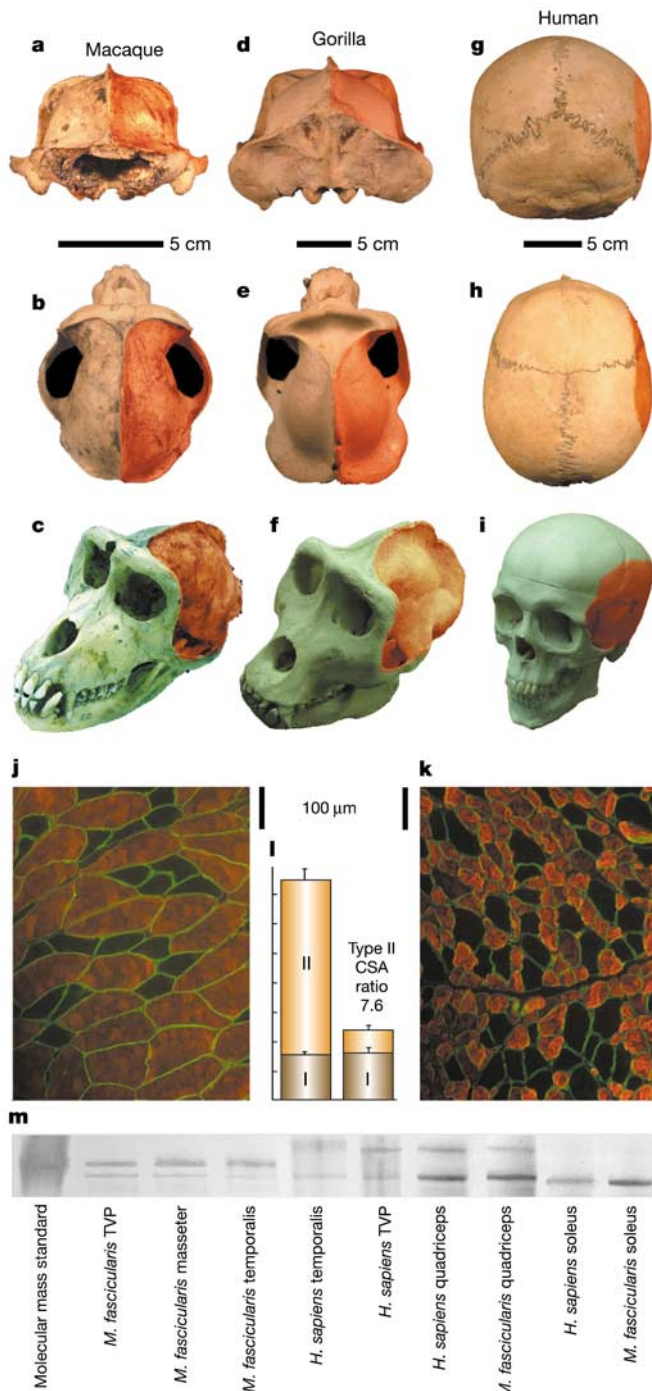


Figure 3 Selective loss of type II myofibre size and muscle bulk in the human temporalis. **a–i**, Occipital, superior and oblique views showing temporal fossae (highlighted in red) and temporalis foramina (circumscribed by the zygomatic arch) in skulls from *M. fascicularis* (**a–c**), *Gorilla gorilla* (**d–f**) and *H. sapiens* (**g–i**). **j, k**, Immunofluorescently stained cryosections of temporalis from *M. fascicularis* (**j**) and *H. sapiens* (**k**). Magnification, $\times 100$. Rhodamine (red) fluorescence indicates type II myofibres, fluorescein (green) reveals interstitial laminin at fibre perimeter. Note that unstained type I fibres are the same size in both species. **l**, Quantification of relative cross-sectional area (CSA) by fibre type. **m**, Coomassie-blue-stained gel electrophoretograms of myosins from limb and first pharyngeal arch muscles of *M. fascicularis* and *H. sapiens*, showing an isomyosin of unique mobility in the first pharyngeal arch derivatives of *M. fascicularis*, midway between MYH1 and -2 bands co-migrating at the top and MYH7 at the bottom¹⁰.

that an abrupt evolutionary alteration in the size and contractile force generated by these muscles would have had pleiotropic effects on craniofacial morphology in the first homozygous *MYH16*-null human ancestor. Such a reduction would probably be inconsistent with the temporalis hypertrophy implicated in sagittal cresting and zygomatic flaring in non-human primates as well as in *Paranthropus* and *Australopithecus*. It is also likely that diminished contractile force would translate into a reduction in stress across patent sutures, sites of dura-mater-patterned growth in the immature neurocranium¹⁵.

To estimate the age of the gene inactivation, we aligned orthologous *MYH16* sequences for human, chimpanzee, orang-utan, macaque and dog, and deduced ancestral sequences (Fig. 4). The high ratio of silent (synonymous) to amino-acid-altering (non-synonymous) mutations indicates that the *MYH16* gene evolved under negative (purifying) darwinian selection in all ancestral lineages except that leading directly to *H. sapiens*¹⁶. The loss of selective constraint after gene inactivation allows non-synonymous mutations to accumulate at the neutral mutation rate¹⁷. Under the assumption that the neutral mutation rate has remained constant since the human–chimpanzee divergence at 6–7 million years (Myr) ago^{17,18}, application of the formula of ref. 17 yields a time of 2.4 ± 0.3 Myr for the appearance of the inactivating mutation in a hominid ancestor (see Supplementary Information).

The earliest hominid fossil record, restricted to the Late Miocene epoch of Africa, is comprised of a variety of taxa featuring small, chimpanzee-sized braincases and large masticatory complexes^{18–21}. Emphasis on a robust masticatory adaptation continues through the Pliocene epoch in various species of *Australopithecus* and *Paranthropus*^{1,22}. In contrast, a relatively gracile masticatory

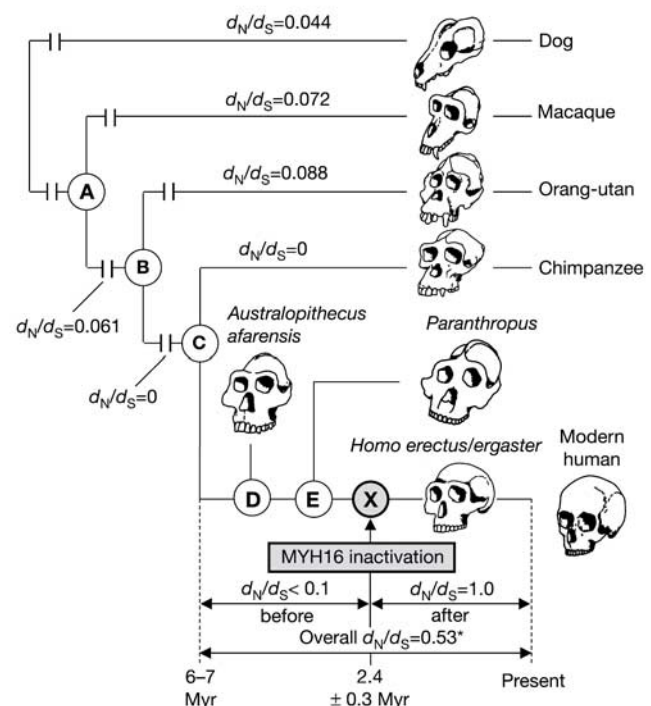


Figure 4 Divergence of the *MYH16* orthologue. Data are based on nucleotide sequences for the six largest exons in the rod-coding domain. Individual mutations are indicated by N for non-synonymous or S for synonymous. d_N and d_S refer to the mutational rates normalized to the number of relevant sites (840 N, 225 S) in pairwise comparisons between *MYH16* orthologues of extant species and inferred ancestors at nodes A, B and C. The atypically high d_N/d_S in the human ancestral lineage (asterisk, $P = 0.0527$) indicates a loss of negative darwinian selection approximately 2.4 Myr ago (see Supplementary Information).

apparatus appears in *Homo erectus/ergaster* by 1.8–2.0 Myr⁵. Coinciding with this loss of masticatory strength is the marked increase in cranial capacity^{4,5}, which characterizes the Pleistocene evolution of *Homo*. Our findings on the age of the inactivating mutation in the *MYH16* gene raise the intriguing possibility that the decrement in masticatory muscle size removed an evolutionary constraint on encephalization, as suggested by the anatomy of the muscle attachments relative to the sutures (compare Fig. 3a and d with Fig. 3g). □

Methods

Bioinformatics

The *MYH16* locus was initially recognized by PCR of human genomic DNA using degenerate primers targeting the coding sequences for two highly conserved α -helices flanking 'loop 1' at the myosin 25/50-kDa junction^{23,24}. As soon as a draft DNA sequence spanning this sequence was posted online (AC112711) a complete representation of the human *MYH16* locus was constructed by integrating this information for the head domain of the gene into the suite of computational algorithms described in ref. 24. The deduced complementary DNA and peptide sequences obtained by restoring an open reading frame across codon 660 (conforming to the primate consensus shown in Fig. 1) were then compared to sequences for all other class II myosin heavy chains encoded by the human genome using the ClustalW alignment algorithm in MacVector 7.1.1 (Accelrys). Phylogenetic reconstruction used the maximum likelihood execution within the program PAUP* 4.0 beta²⁵. All other methods, as detailed in Supplementary Information, gave a topologically similar tree with this input data set. Ancestral sequence reconstruction was achieved independently by using the maximum parsimony (PAMP) and maximum likelihood (BASEML) programs within the most recent release of the program PAML²⁶; the results were identical, reflecting the degree of conservation among the aligned sequences. The program suite MEGA version 2.1 (ref. 27) was used to calculate d_N/d_S (see Fig. 4 legend for definition) ratios and relevant statistics on a per lineage basis.

Messenger RNA isolation, PCR and sequence analysis

Messenger RNA was isolated using the MicroFastTrack 2.0 mRNA isolation kit (Invitrogen). Complementary DNA was synthesized using the Retroscript first strand synthesis kit (Ambion). *MYH16* RT-PCR amplifications were performed using HotStarTaq DNA Polymerase (Qiagen). Genomic DNA samples obtained from the Coriell Genetic Cell Repositories and the Southwest Foundation National Primate Research Center were amplified by PCR using the primer pairs and annealing temperatures listed in Supplementary Information. After purification of PCR products using Qiagen QIAquick gel extraction kit, according to the manufacturer's specifications, bidirectional DNA sequences were generated by the University of Pennsylvania School of Medicine, Department of Genetics Sequencing Core, after fluorimetric analysis of dideoxynucleotide chain termination products.

Protein gel electrophoresis and peptide analysis

Myofibrils from human and cynomolgus macaque tissues were isolated and washed as described²⁸. Polyacrylamide gel electrophoresis was performed as described²⁹, using 4 μ g of total protein per lane. Myosin isoforms were visualized by staining gels with Bio-safe Coomassie blue stain (BioRad). A gel slice containing the dominant myosin heavy chain isoform from the *M. fascicularis* temporalis muscle was subjected to trypsin proteolysis and tandem mass spectrometry peptide sequence analysis at the Wistar Institute Proteomics Facility. The analysis was performed using microcapillary reverse-phase high-performance liquid chromatography nano-spray tandem mass spectrometry on a ThermoFinnigan LCQ quadrupole ion-trap mass spectrometer. The tandem mass spectrometry spectra were run against the non-redundant human protein sequence database and the predicted products of the newly recognized *MYH14*, -15 and -16 genes using the SEQUEST software package³⁰.

Tissue procurement

All human muscle samples were obtained under protocol number 203200 as approved by the Hospital of the University of Pennsylvania Institutional Review Board. All primate tissues were obtained under protocol number 706976 as approved by the University of Pennsylvania Animal Care and Use Committee.

Antibody staining protocol

Frozen tissues were sectioned at a thickness of 8 μ m on a Leica cryotome. Sections were blocked with 2% filter-sterilized bovine serum albumin (BSA) for 3.5 h. BSA solution was removed without washing and replaced with a solution containing mouse-derived primary antibody against 'Fast' (Sigma, M-4276; dilution 1:800), 'Slow' (Sigma, M-8421; dilution 1:5,000), or IIA (SC-71; dilution 1:100) myosin heavy chain protein, as well as rabbit-derived primary antibody against laminin (Sigma, L-9393; dilution 1:2,500), and incubated overnight at 4 °C. Sections were then washed three times with 1 $\times$ PBS and incubated for 1 h in a secondary antibody solution containing Cy3-conjugated goat anti-mouse IgG (Jackson ImmunoResearch, 113-165-100; dilution 1:200) and fluorescein-isothiocyanate-conjugated sheep anti-rabbit IgG (Sigma, F-7512; dilution 1:160). After secondary antibody incubation, sections were washed three times in 1 $\times$ PBS and mounted with Vectashield containing 4,6-diamidino-2-phenylindole. Photographs were taken at $\times$ 100 total magnification using a Zeiss Axiophot2 microscope equipped with a Nikon DXM1200 digital camera.

Image analysis

Fibre size was determined using Image Pro Plus (Phase 3 Imaging Systems), running a macro optimized to calculate fibre area based on a laminin stain.

Received 5 April 2003; accepted 20 January 2004; doi:10.1038/nature02358.

1. Rak, Y. *The Australopithecine Face* (Academic, New York, 1983).
2. Aiello, L. & Dean, C. *An Introduction to Human Evolutionary Anatomy* (Academic, New York, 1998).
3. White, T. D., Suwa, G., Simpson, S. & Asfaw, B. Jaws and teeth of *Australopithecus afarensis* from Maka, Middle Awash, Ethiopia. *Am. J. Phys. Anthropol.* **111**, 45–68 (2000).
4. Tobias, P. V. *The Skulls, Endocrania, and Teeth of Homo habilis* (University Press, Cambridge, 1991).
5. Walker, A. & Leakey, R. (eds) *The Nariokotome Homo erectus skeleton* (Harvard Univ. Press, Cambridge, Massachusetts, 1993).
6. Vekua, A. *et al.* A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297**, 85–89 (2002).
7. Acakpo-Satchivi, L. *et al.* Growth and muscle defects in mice lacking adult myosin heavy chain genes. *J. Cell Biol.* **139**, 1219–1229 (1997).
8. Allen, D. L., Harrison, B. C., Sartorius, C., Byrnes, W. C. & Leinwand, L. A. Mutation of the IIB myosin heavy chain gene results in muscle fiber loss and compensatory hypertrophy. *Am. J. Physiol. Cell Physiol.* **280**, C637–C645 (2001).
9. Martinsson, T. *et al.* Autosomal dominant myopathy: missense mutation (Glu-706 $\rightarrow$ Lys) in the myosin heavy chain IIa gene. *Proc. Natl Acad. Sci. USA* **97**, 14614–14619 (2000).
10. Korfage, J. A. & Van Eijden, T. M. Myosin heavy chain composition in human masticatory muscles by immunohistochemistry and gel electrophoresis. *J. Histochem. Cytochem.* **51**, 113–119 (2003).
11. McPherron, A. C., Lawler, A. M. & Lee, S. J. Regulation of skeletal muscle mass in mice by a new TGF- β superfamily member. *Nature* **387**, 83–90 (1997).
12. Hamrick, M. W., McPherron, A. C., Lovejoy, C. O. & Hudson, J. Femoral morphology and cross-sectional geometry of adult myostatin-deficient mice. *Bone* **27**, 343–349 (2000).
13. Hohl, T. H. Masticatory muscle transposition in primates: effects on craniofacial growth. *J. Maxillofac. Surg.* **11**, 149–156 (1983).
14. Brennan, M. & Antonyshyn, O. The effects of temporalis muscle manipulation on skull growth: an experimental study. *Plast. Reconstr. Surg.* **97**, 13–24 (1996).
15. Warren, S. M., Brunet, L. J., Harland, R. M., Economides, A. N. & Longaker, M. T. The BMP antagonist noggin regulates cranial suture fusion. *Nature* **422**, 625–629 (2003).
16. Messier, W. & Stewart, C. B. Episodic adaptive evolution of primate lysozymes. *Nature* **385**, 151–154 (1997).
17. Chou, H. H. *et al.* Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl Acad. Sci. USA* **99**, 11736–11741 (2002).
18. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
19. Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).
20. Leakey, M. G., Feibel, C. S., McDougall, I., Ward, C. & Walker, A. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature* **393**, 62–66 (1998).
21. Lockwood, C. A., Kimbel, W. H. & Johanson, D. C. Temporal trends and metric variation in the mandibles and dentition of *Australopithecus afarensis*. *J. Hum. Evol.* **39**, 23–55 (2000).
22. Asfaw, B. *et al.* *Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science* **284**, 629–635 (1999).
23. Uyeda, T. Q. P., Ruppel, K. M. & Spudich, J. A. Enzymatic activities correlate with chimaeric substitutions at the actin-binding face of myosin. *Nature* **368**, 567–569 (1994).
24. Desjardins, P., Burkman, J., Shrager, J., Allmond, L. & Stedman, H. Evolutionary implications of three novel members of the human sarcomeric myosin heavy chain gene family. *Mol. Biol. Evol.* **19**, 375–393 (2002).
25. Swofford, D. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer Associates, Sunderland, Massachusetts, 1998).
26. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**, 555–556 (1997).
27. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245 (2001).
28. Bamman, M. M., Clarke, M. S., Talmadge, R. J. & Feeback, D. L. Enhanced protein electrophoresis technique for separating human skeletal muscle myosin heavy chain isoforms. *Electrophoresis* **20**, 466–468 (1999).
29. Levine, S., Kaiser, L., Leferovich, J. & Tikunov, B. Cellular adaptations in the diaphragm in chronic obstructive pulmonary disease. *N. Engl. J. Med.* **337**, 1799–1806 (1997).
30. MacCoss, M. J., Wu, C. C. & Yates, J. R. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593–5599 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Brayman, N. Mirza, M. Ruckenstein and the University of Washington National Primate Research Center (Seattle) for providing access to the biopsy material used in this study; N. Gilmore of the Philadelphia Academy of Natural Sciences for access to specimens for photography; L. Joseph, D. Fonseca, R. McCourt and W. Ewens for assistance with the bioinformatic analysis; and P. Dodson, L. Whitaker, A. Kelly and S. Bartlett for advance reading of the manuscript. We also thank colleagues in the University of Pennsylvania Genomics Institute Bioinformatics Core, and Wistar Institute Proteomics Facility for their assistance. This work was supported in part by grants to H.H.S. from the NIH (NIAMS and NINDS), MDA, AFM, VA and the Genzyme Corporation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.H.S. (hstedman@mail.med.upenn.edu).

General conditions for predictivity in learning theory

Tomaso Poggio¹, Ryan Rifkin^{1,4}, Sayan Mukherjee^{1,3} & Partha Niyogi²

¹Center for Biological and Computational Learning, McGovern Institute Computer Science Artificial Intelligence Laboratory, Brain Sciences Department, MIT, Cambridge, Massachusetts 02139, USA

²Departments of Computer Science and Statistics, University of Chicago, Chicago, Illinois 60637, USA

³Cancer Genomics Group, Center for Genome Research/Whitehead Institute, MIT, Cambridge, Massachusetts 02139, USA

⁴Honda Research Institute USA Inc., Boston, Massachusetts 02111, USA

Developing theoretical foundations for learning is a key step towards understanding intelligence. 'Learning from examples' is a paradigm in which systems (natural or artificial) learn a functional relationship from a training set of examples. Within this paradigm, a learning algorithm is a map from the space of training sets to the hypothesis space of possible functional solutions. A central question for the theory is to determine conditions under which a learning algorithm will generalize from its finite training set to novel examples. A milestone in learning theory¹⁻⁵ was a characterization of conditions on the hypothesis space that ensure generalization for the natural class of empirical risk minimization (ERM) learning algorithms that are based on minimizing the error on the training set. Here we provide conditions for generalization in terms of a precise stability property of the learning process: when the training set is perturbed by deleting one example, the learned hypothesis does not change much. This stability property stipulates conditions on the learning map rather than on the hypothesis space, subsumes the classical theory for ERM algorithms, and is applicable to more general algorithms. The surprising connection between stability and predictivity has implications for the foundations of learning theory and for the design of novel algorithms, and provides insights into problems as diverse as language learning and inverse problems in physics and engineering.

One of the main impacts of learning theory is on engineering. Systems that learn from examples to perform a specific task have many applications⁶. For instance, a system may be needed to recognize whether an image contains a face or not. Such a system could be trained with positive and negative examples: images with and without faces. In this case, the input image is a point in a multidimensional space of variables such as pixel values; its associated output is a binary 'yes' or 'no' label.

In the auditory domain, one may consider a variety of problems. Consider speaker authentication. The input is an acoustic utterance, and the system has to determine whether it was produced by a particular target speaker or not. Training examples would then consist of a set of utterances each labelled according to whether or not they were produced by the target speaker. Similarly, in speech recognition, one wishes to learn a function that maps acoustic utterances to their underlying phonetic sequences. In learning the syntax of a language, one wishes to learn a function that maps sequences of words to their grammaticality values. These functions could be acquired from training data.

In another application in computational biology, algorithms have been developed that can produce a diagnosis of the type of cancer from a set of measurements of the expression level of many thousands of human genes in a biopsy of the tumour measured with a complementary DNA microarray containing probes for a number of genes. Again, the software learns the classification rule from a set of examples, that is, from examples of expression patterns in a number of patients with known diagnoses.

What we assume in the above examples is a machine that is trained, instead of programmed, to perform a task, given data of the form $S = (x_i, y_i)_{i=1}^n$. Training means synthesizing a function that best represents the relation between the inputs x_i and the corresponding outputs y_i .

The basic requirement for any learning algorithm is generalization: the performance on the training examples (empirical error) must be a good indicator of the performance on future examples (expected error), that is, the difference between the two must be 'small' (see Box 1 for definitions; see also Fig. 1).

Probably the most natural learning algorithm is ERM: the algorithm 'looks' at the training set S , and selects as the estimated function the one that minimizes the empirical error (training error) over the functions contained in a hypothesis space of candidate

Box 1

Formal definitions in supervised learning

Convergence in probability. A sequence of random variables $\{X_n\}$ converges in probability to a random variable X (for example, $\lim_{n \rightarrow \infty} |X_n - X| = 0$ in probability) if and only if for every $\epsilon > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$.

Training data. The training data comprise input and output pairs. The input data X is assumed to be a compact domain in an euclidean space and the output data Y is assumed to be a closed subset of $\mathbb{R}^k$. There is an unknown probability distribution $\mu(x, y)$ on the product space $Z = X \times Y$. The training set S consists of n independent and identically drawn samples from the distribution on Z :

$$S = \{z_1 = (x_1, y_1), \dots, z_n = (x_n, y_n)\}$$

Learning algorithms. A learning algorithm takes as input a data set S and outputs a function f_S that represents the relation between the input x and output y . Formally the algorithm can be stated as a map $L : \cup_{n \geq 1} Z^n \rightarrow \mathcal{H}$ where $\mathcal{H}$, called the hypothesis space, is the space of functions the algorithm 'searches' to select f_S . We assume that the algorithm is symmetric, that is, f_S does not depend on the ordering of the samples in S . Most learning algorithms are either regression or classification algorithms depending on whether y is real-valued or binary valued.

Loss functions. We denote the price we pay with $V(f, z)$ when the prediction for a given x is $f(x)$ and the true value is y . We assume that $V(f, z)$ is always bounded. A classical example of a loss function is the square loss $V(f, z) = (f(x) - y)^2$.

Expected error. The expected error of a function f is defined as

$$I[f] = \int_Z V(f, z) d\mu(z)$$

which is also the expected error of a new sample z drawn from the distribution. In the case of square loss:

$$I[f] = \int_{X \times Y} (f(x) - y)^2 d\mu(x, y)$$

We would like to find functions for which $I[f]$ is small. However, we cannot compute $I[f]$ because we are not given the distribution μ .

Empirical error. The following quantity, called empirical error, can be computed given the training data S :

$$I_S[f] = \frac{1}{n} \sum_{i=1}^n V(f, z_i)$$

Generalization and consistency. An algorithm generalizes if the function f_S selected by it satisfies for all S ($|S| = n$) and uniformly for any probability distribution μ

$$\lim_{n \rightarrow \infty} |I[f_S] - I_S[f_S]| = 0 \text{ in probability}$$

An algorithm is (universally) consistent if uniformly for any distribution μ and any $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ I[f_S] > \inf_{f \in \mathcal{H}} I[f] + \epsilon \right\} = 0$$

functions. Classical learning theory was developed around the study of ERM. One of its main achievements is a complete characterization of the necessary and sufficient conditions for generalization of ERM and its consistency^{1–5}. Consistency (see Box 2) requires that the expected error of the solution converges to the expected error of the most accurate function in the hypothesis class $\mathcal{H}$. For ERM, generalization is equivalent to consistency³. Generalization of ERM can be ensured by restricting sufficiently the hypothesis space $\mathcal{H}$ (ref. 7 provides a theory within a classical mathematical framework).

The basic intuition here is that if the class $\mathcal{H}$ is too large, in the sense of containing too many wild functions, it is impossible to construct any useful approximating function using empirical data. Without restrictions on $\mathcal{H}$ there are functions that minimize the empirical error by fitting the data S exactly (thus $I_S[f_S] = 0$) but are very far away from the ‘true’ underlying function and therefore have a large expected error (thus $I[f_S]$ is large).

Although the classical theory has achieved a complete characterization of ERM, there are many learning algorithms, some existing and some likely to be discovered, that are not ERM. Several of the most effective, state-of-the-art algorithms—from square-loss regularization⁸ to support vector machines², from bagging⁹ to boosting^{10, 11}, from k -nearest neighbour¹² to vicinal risk minimization¹³—are not, strictly, ERM, because ERM is defined as minimization of the empirical error (see Box 1) within a fixed—a priori—hypothesis space. Though specific theoretical results can be derived for some of these algorithms^{7, 14, 15}, general conditions, representing a broadly applicable approach for checking generalization properties of any learning algorithm, would be desirable. For the case of ERM algorithms, Vapnik and Červonenkis² asked: what property must the hypothesis space $\mathcal{H}$ have for good generalization of ERM? For the case of more general algorithms, we note that any learning algorithm is a map L from data sets to hypothesis functions. For a general theory, it therefore makes sense to ask: what property must the learning map L have for good generalization error? Ideally a general answer to the second question must include an answer to the first question as well.

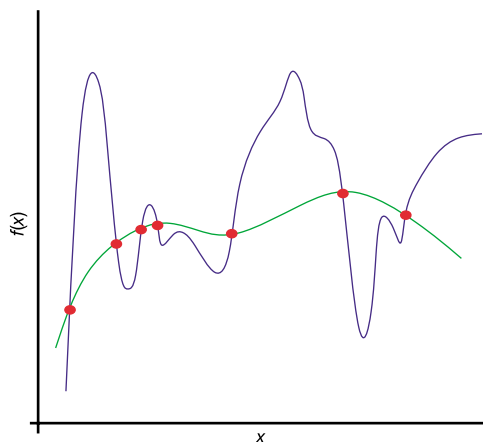


Figure 1 Example of an empirical minimizer with large expected error. In the case sketched here the data were generated from the ‘true’ green function. The blue function fits the data set and therefore has zero empirical error ($I_S[f_{\text{blue}}] = 0$). Yet it is clear that on future data, this function f_{blue} will perform poorly as it is far from the true function on most of the domain. Therefore $I[f_{\text{blue}}]$ is large. Generalization refers to whether the empirical performance on the training set ($I_S[f]$) will generalize to test performance on future examples ($I[f]$). If an algorithm is guaranteed to generalize, an absolute measure of its future predictivity can be determined from its empirical performance.

Recently, Bousquet and Elisseeff¹⁵ proposed the notion of uniform stability to characterize the generalization properties of an algorithm (see Box 3). The idea—closely related to stability of well-posed problems¹⁶—is to look at the stability of the learning map L by considering how much the error at a point changes if the training set is perturbed. Uniform stability was too strong a stability requirement to subsume the classical theory (see Box 3), so the search remained open. Many different notions of algorithmic stability exist, going back at least to refs 14 and 17.

We first introduce the definition of cross-validation leave-one-out (CV_{loo}) stability: the learning map L is distribution-independent, CV_{loo} stable if uniformly for all probability distributions μ $\lim_{n \rightarrow \infty} \sup_{i \in \{1, \dots, n\}} |V(f_{S^i}, z_i) - V(f_S, z_i)| = 0$ in probability, where S^i denotes the training set S with the i th point removed.

CV_{loo} stability measures the difference in errors at a point z_i between a function obtained given the entire training set and one obtained given the same training set but with the point z_i left out (see Fig. 2). CV_{loo} stability is strictly weaker than uniform stability because the condition holds only on most (note the probabilistic quantification) training points and not for all possible points z (see also Box 3). For the supervised learning setting, the definition of CV_{loo} stability implements a specific and weak form of the general idea of stability of a well-posed problem: the function ‘learned’ from a training set should, with high probability, change little in its predictions for a small change in the training set, such as deletion of one of the examples.

The first crucial question is whether CV_{loo} stability is general enough to subsume the classical conditions of the consistency of ERM. The answer is surprising and positive¹⁶:

Theorem A. For ‘good’ loss functions the following statements are equivalent for ERM: (1) L is distribution-independent CV_{loo} stable; (2) ERM generalizes and is universally consistent; (3) $\mathcal{H}$ is uniform Glivenko–Cantelli (uGC; see Box 2).

Box 2

Classical results of empirical risk minimization algorithms

Empirical risk minimization (ERM) algorithms are defined as those satisfying:

$$I_S[f_S] = \min_{f \in \mathcal{H}} I_S[f]$$

The results described in this Letter for the special case of exact minimization are also valid in the general case of almost ERM, in which the minimum is not assumed to exist (see ref. 16), though the proofs are technically somewhat more complex.

The key theorem^{1–5} of classical learning theory relates consistency of ERM to a constraint on the function classes $\mathcal{H}^2$.

Theorem. The following are equivalent for ‘well-behaved’ loss functions, such as the square-loss; (1) ERM generalizes and is consistent; (2) $\mathcal{H}$ is a uGC class.

A function class is a uGC class if universal, uniform convergence in probability holds

$$\lim_{n \rightarrow \infty} \sup_{\mu} \mathbb{P} \left(\sup_{f \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n f(x_i) - \int_{\mathcal{X}} f(x) d\mu(x) \right| > \varepsilon \right) = 0.$$

The result extends to general loss functions if the functions induced by the composition of the loss V and hypothesis space $\mathcal{H}$ are uGC. For binary functions the uGC property reduces to the well known requirement of finite VC dimension.

Cucker and Smale⁷ and Zhou²⁵ developed a complete and effective theory exploiting the property of compactness of $\mathcal{H}$, which is sufficient, but not necessary, for generalization and consistency of ERM, because compactness of $\mathcal{H}$ implies the uGC property of $\mathcal{H}$ but not vice versa.

We now ask whether CV_{100} stability is sufficient for generalization of any learning algorithm satisfying it. The answer to this question is negative (Theorem 11 in ref. 15 claims incorrectly that CV_{100} stability is sufficient for generalization, since there are counter-examples¹⁶).

A positive answer can be obtained by augmenting CV_{100} stability with stability of the expected error and stability of the empirical error to define a new notion of stability, CV_{100} stability, as follows: the learning map L is distribution-independent, CV_{100} stable if for all probability distributions μ : (1) is CV_{100} stable; (2) $\lim_{n \rightarrow \infty} \sup_{i \in \{1, \dots, n\}} |I[f_S] - I[f_{S'}]| = 0$ in probability; (3) $\lim_{n \rightarrow \infty} \sup_{i \in \{1, \dots, n\}} |I_S[f_S] - I_S[f_{S'}]| = 0$ in probability.

Properties (2) and (3) are weak and satisfied by most 'reasonable' learning algorithms. They are not sufficient for generalization. In the case of ERM CV_{100} stability is the key property since both conditions (2) and (3) are implied by consistency of ERM¹⁶. Unlike CV_{100} stability, CV_{100} stability is sufficient for generalization for any learning algorithm¹⁶.

Theorem B. If a learning algorithm is CV_{100} stable and the loss function is bounded, then f_S generalizes, that is uniformly for all μ $\lim_{n \rightarrow \infty} |I[f_S] - I_S[f_S]| = 0$ in probability.

Notice that theorem B provides a general condition for generalization but not for consistency, which is not too surprising because the class of non-ERM algorithms is very large. In summary, CV_{100} stability is sufficient for generalization for any algorithm and necessary and sufficient for generalization and consistency of ERM.

Good generalization ability means that the performance of the algorithm on the training set will accurately reflect its future performance on the test set. Therefore an algorithm that guarantees good generalization will predict well if its empirical error on the training set is small. Conversely, notice that it is also possible for an algorithm to generalize well but predict poorly (when both empirical and expected performances are poor). Crucially, therefore, one can empirically determine the predictive performance by looking at the error on the training set.

Box 3 Uniform stability

There have been many different notions of stability going back of Tikhonov²⁶, Devroye and Wagner²⁷ and Kearns and Ron²⁸. A significant step was taken recently by Bousquet and Elisseeff¹⁵, who defined uniform stability. The learning map L is uniformly stable if

$$\forall S \in Z^n, \forall i \in \{1, \dots, n\} \sup_{z \in Z} |V(f_S, z) - V(f_{S'}, z)| \leq \beta^{(n)}$$

and $\beta^{(n)} = K/n$, where K is a constant. Uniform stability is a natural measure of continuity of L because the supremum measures the largest change at any point z . Uniform stability implies good generalization¹⁵, and Tikhonov regularization algorithms (including support vector machines²⁹) are uniformly stable¹⁵.

Unfortunately, uniform stability is a very strong requirement because the change in error when a point is removed must be small for any $x, y \in Z$ and any training set. Most algorithms are not uniformly stable. For example, ERM with a hypothesis space of only two functions is not guaranteed to be uniformly stable¹⁷.

The search remained open for a notion of stability that is sufficient for generalization, and necessary and sufficient for ERM. A partial answer was provided by Kutin and Niyogi¹⁷ when they introduced the notion of cross validation (CV) stability and showed that this (with an additional weaker condition) was adequate in the classical Probably Approximately Correct (PAC) setting³⁰. A general answer was not found.

In this Letter, we show that a leave-one-out version of CV stability (CV_{100}) along with some more technical conditions provides the answer.

Learning techniques are similar to fitting a multivariate function to measurement data, a classical problem in nonparametric statistics^{10,18,19}. The key point is that the fitting should be predictive. In this sense 'learning from examples' can also be considered as a stylized model for the scientific process of developing predictive theories from empirical data². The classical conditions for generalization of ERM can then be regarded as the formalization of a 'folk theorem', which says that simple theories should be preferred among all of those that fit the data. Our stability conditions would instead correspond to the statement that the process of research should—most of the time—only incrementally change existing scientific theories as new data become available.

It is intellectually pleasing that the concept of stability, which cuts across so many different areas of mathematics, physics and engineering, turns out to play such a key role in learning theory. It is somewhat intuitive that stable solutions are predictive, but it is especially surprising that our specific definition of CV_{100} stability fully subsumes the classical necessary and sufficient conditions on $\mathcal{H}$ for consistency of ERM.

In this Letter we have stated the main properties in an asymptotic form. The detailed statements¹⁶ of the two stability theorems sketched here, however, provide bounds on the difference between empirical and expected error for any finite size of the training set.

Our observations on stability suggest immediately several other questions. The first challenge is to bridge learning theory and a quite different and broad research area—the study of inverse problems in applied mathematics and engineering²⁰—since stability is a key condition for the solution of inverse problems (our stability condition can be seen as an extension¹⁶ to the general learning problem of the classical notion of condition number that characterize stability of linear systems). In fact, while predictivity is at the core of classical learning theory, another motivation drove the development of several of the best existing algorithms (such as regularization algorithms of which SVMs are a special case): well-

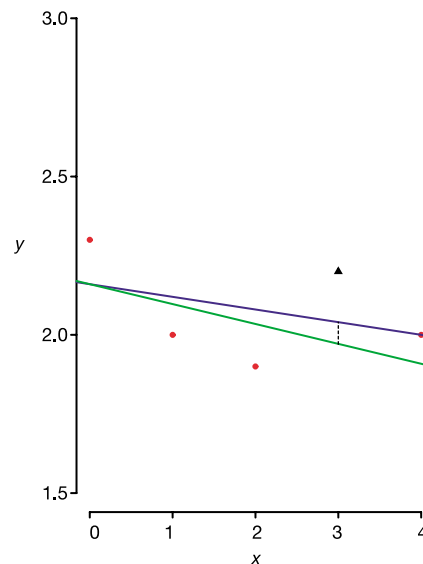


Figure 2 Measuring CV_{100} stability in a simple case. The blue line was obtained by a linear regression (for example, ERM with square loss on a hypothesis space of linear functions) on all five training points ($n = 5$). The green line was obtained in the same way by 'leaving out' the black triangle from the training set. In this case, CV_{100} stability requires that when a single point is removed from a data set, the change in error at the removed point (here indicated by the black dashed line) is small and decreases to zero in probability for n increasing to infinity.

posedness and, specifically, stability of the solution. These two requirements—consistency and stability—have been treated so far as ‘defacto’ separate, and in fact there was no a priori reason to believe that they are related⁶. Our new result shows these two apparently different motivations are actually completely equivalent for ERM.

The most immediate implication of $\text{CVEEE}_{\text{loo}}$ stability and its properties has to do with developing learning theory beyond the ERM approach. In particular, $\text{CVEEE}_{\text{loo}}$ stability can provide generalization bounds for algorithms other than ERM. For some of them a ‘VC-style’ analysis (see Box 2) in terms of complexity of the hypothesis space can still be used; for others, such as k -nearest neighbour, such an analysis is impossible because the hypothesis space has unbounded complexity or is not even defined, whereas $\text{CVEEE}_{\text{loo}}$ stability can still be used. Though a detailed analysis for specific algorithms needs to be done, some interesting observations in terms of stability can be inferred easily from existing analyses. For instance, the results of ref. 15 imply that regularization and SVMs are $\text{CVEEE}_{\text{loo}}$ stable; a version of bagging with the number k of regressors increasing with n is $\text{CVEEE}_{\text{loo}}$ stable²¹; k -nearest neighbour with $k \rightarrow \infty$ and $k/n \rightarrow 0$ and kernel rules with the width $h_n \rightarrow 0$ and $h_n n \rightarrow \infty$ are also $\text{CVEEE}_{\text{loo}}$ stable. Thus because of theorem B, all these algorithms have the generalization property (and some are also universally consistent).

More importantly, $\text{CVEEE}_{\text{loo}}$ stability may also suggest new algorithms that, unlike ERM, enforce stability directly. Similarly, it may be possible to gain a better understanding of existing statistical tests and develop new ones based on the definition of $\text{CVEEE}_{\text{loo}}$ stability and extensions of it. Furthermore, the search for equivalent and possibly ‘simpler’ conditions than $\text{CVEEE}_{\text{loo}}$ stability is open.

A theory of learning based on stability may have more direct connections with cognitive properties of the brain’s mechanisms for learning (needless to say, learning is much more than memory). Stability is a condition on the learning machinery (or algorithms), as the classical conditions, such as the uGC property of $\mathcal{H}$, constrain the domain of learning—the hypothesis space. It is interesting that neural circuits approximating algorithms such as radial basis functions have been proposed as learning modules in the brain^{22,23}. Such algorithms are $\text{CVEEE}_{\text{loo}}$ stable because they follow from regularization. There are certainly other learning algorithms that are biologically plausible, do not follow from regularization or ERM, but may be analysable in terms of $\text{CVEEE}_{\text{loo}}$ stability.

Consider for example, the problem of learning a language. The language learning algorithm $\mathcal{A}_L$ is a map from linguistic data (sentences produced by people) to computable functions (grammars) that are learned from those data. Corresponding to the learning algorithm $\mathcal{A}_L$ there exists a class $H_{\mathcal{A}_L}$ which is the class of all learnable grammars. Thus $H_{\mathcal{A}_L}$ is the hypothesis class corresponding to the language learning algorithm. In the tradition of generative linguistics identified most strongly with Chomsky, the class of possible natural language grammars $H_{\mathcal{A}_L}$ is called ‘universal grammar’, and different linguistic theories attempt to characterize the nature of this class. For example, the principles and parameters approach²⁴ tries to describe this class in a parametric fashion with a finite number of boolean parameters.

Although this tradition in generative linguistics is a meaningful one, it reflects an emphasis on the hypothesis class. In this sense, it is in the same philosophical spirit as the classical approach to learning theory of Vapnik and Červonenkis where conditions on the hypothesis space (expressed in terms of the so called VC dimension) are outlined for learnability. While $\mathcal{A}$ and $\mathcal{H}_{\mathcal{A}}$ are related, it is possible that in many cases, $\mathcal{A}$ may admit an easier mathematical characterization than $\mathcal{H}_{\mathcal{A}}$. Thus, for example, it may be possible that the language learning algorithm may be easy to describe mathematically while the class of possible natural language grammars may be difficult to describe. In that case, the shift in focus to the algorithm

rather than the hypothesis class would correspond to a shift to a stability rather than a VC point of view.

It is often the case in the natural world that multiple representations of the same underlying phenomena are formally equivalent but some representations are more insightful than others. In the case of natural language, only time will tell whether $\mathcal{A}$ or $\mathcal{H}_{\mathcal{A}}$ will prove to be the easier and more insightful object. At present, learning theory in the context of language focuses heavily on the latter, and many ‘learnable classes’ are studied and theories are developed about them. We hope that focus on the learning algorithm may stimulate a new kind of language learning theory and practice. $\square$

Received 3 November 2003; accepted 16 January 2004; doi:10.1038/nature02341.

1. Vapnik, V. & Červonenkis, A. Y. The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognit. Image Anal.* **1**, 283–305 (1991).
2. Vapnik, V. N. *Statistical Learning Theory* (Wiley, New York, 1998).
3. Alon, N., Ben-David, S., Cesa-Bianchi, N. & Haussler, D. Scale-sensitive dimensions, uniform convergence, and learnability. *J. Assoc. Comp. Mach.* **44**, 615–631 (1997).
4. Dudley, R. M. *Uniform Central Limit Theorems* (Cambridge studies in advanced mathematics, Cambridge Univ. Press, 1999).
5. Dudley, R., Gine, E. & Zinn, J. Uniform and universal Glivenko–Cantelli classes. *J. Theor. Prob.* **4**, 485–510 (1991).
6. Poggio, T. & Smale, S. The mathematics of learning: Dealing with data. *Not. Am. Math. Soc.* **50**, 537–544 (2003).
7. Cucker, F. & Smale, S. On the mathematical foundations of learning. *Bull. Am. Math. Soc.* **39**, 1–49 (2001).
8. Wahba, G. *Spline Models for Observational Data* (Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990).
9. Breiman, L. Bagging predictors. *Machine Learn.* **24**, 123–140 (1996).
10. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer series in statistics, Springer, Basel, 2001).
11. Freund, Y. & Schapire, R. A decision-theoretic generalization of online learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997).
12. Fix, E. & Hodges, J. Discriminatory analysis, nonparametric discrimination: consistency properties? (Techn. rep. 4, Project no. 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX, 1951).
13. Bottou, L. & Vapnik, V. Local learning algorithms. *Neural Comput.* **4**(6), 888–900 (1992).
14. Devroye, L. & Wagner, T. Distribution-free performance bounds for potential function rules. *IEEE Trans. Inform. Theory* **25**, 601–604 (1979).
15. Bousquet, O. & Elisseeff, A. Stability and generalization. *J. Machine Learn. Res.* **2**, 499–526 (2001).
16. Mukherjee, S., Niyogi, P., Poggio, T. & Rifkin, R. Statistical learning: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization (CBCL Paper 223, Massachusetts Institute of Technology, 2002, revised 2003).
17. Kutin, S. & Niyogi, P. in *Proceedings of Uncertainty in AI* (eds Daruich, A. & Friedman, N.) (Morgan Kaufmann, Univ. Alberta, Edmonton, 2002).
18. Stone, C. The dimensionality reduction principle for generalized additive models. *Ann. Stat.* **14**, 590–606 (1986).
19. Donoho, D. & Johnstone, I. Projection-based approximation and a duality with kernel methods. *Ann. Stat.* **17**, 58–106 (1989).
20. Engl, H., Hanke, M. & Neubauer, A. *Regularization of Inverse Problems* (Kluwer Academic, Dordrecht, 1996).
21. Evgeniou, T., Pontil, M. & Elisseeff, A. Leave one out error, stability, and generalization of voting combinations of classifiers. *Machine Learn.* (in the press).
22. Pouget, A. & Sejnowski, T. J. Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* **9**, 222–237 (1997).
23. Poggio, T. A theory of how the brain might work. *Cold Spring Harbor Symp. Quant. Biol.* **55**, 899–910 (1990).
24. Chomsky, N. *Lectures on Government and Binding* (Foris, Dordrecht, 1995).
25. Zhou, D. The covering number in learning theory. *J. Complex.* **18**, 739–767 (2002).
26. Tikhonov, A. N. & Arsenin, V. Y. *Solutions of Ill-posed Problems* (Winston, Washington DC, 1977).
27. Devroye, L. & Wagner, T. Distribution-free performance bounds for potential function rules. *IEEE Trans. Inform. Theory* **25**, 601–604 (1979).
28. Kearns, M. & Ron, D. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural Comput.* **11**, 1427–1453 (1999).
29. Evgeniou, T., Pontil, M. & Poggio, T. Regularization networks and support vector machines. *Adv. Comput. Math.* **13**, 1–50 (2000).
30. Valiant, L. A theory of the learnable. *Commun. Assoc. Comp. Mach.* **27**, 1134–1142 (1984).

Acknowledgements We thank D. Panchenko, R. Dudley, S. Mendelson, A. Rakhlin, F. Cucker, D. Zhou, A. Verri, T. Evgeniou, M. Pontil, P. Tamayo, M. Poggio, M. Calder, C. Koch, N. Cesa-Bianchi, A. Elisseeff, G. Lugosi and especially S. Smale for several insightful and helpful comments. This research was sponsored by grants from the Office of Naval Research, DARPA and National Science Foundation. Additional support was provided by the Eastman Kodak Company, Daimler-Chrysler, Honda Research Institute, NEC Fund, NTT, Siemens Corporate Research, Toyota, Sony and the McDermott chair (T.P.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.P. (tp@ai.mit.edu).

Imaging cortical correlates of illusion in early visual cortex

Dirk Jancke, Frédéric Chavane, Shmuel Naaman & Amiram Grinvald

Department of Neurobiology and the Grodetsky Center for Studies of Higher Brain Function, The Weizmann Institute of Science, 76100 Rehovot, Israel

Exploring visual illusions reveals fundamental principles of cortical processing. Illusory motion perception of non-moving stimuli was described almost a century ago by Gestalt psychologists^{1,2}. However, the underlying neuronal mechanisms remain unknown. To explore cortical mechanisms underlying the 'line-motion' illusion³, we used real-time optical imaging^{4–6}, which is highly sensitive to subthreshold activity. We examined, in the visual cortex of the anaesthetized cat, responses to five stimuli: a stationary small square and a long bar; a moving square; a drawn-out bar; and the well-known line-motion illusion³, a stationary square briefly preceding a long stationary bar presentation. Whereas flashing the bar alone evoked the expected localized,

short latency and high amplitude activity patterns^{7,8}, presenting a square 60–100 ms before a bar induced the dynamic activity patterns resembling that of fast movement. The preceding square, even though physically non-moving, created gradually propagating subthreshold cortical activity that must contribute to illusory motion, because it was indistinguishable from cortical representations of real motion in this area. These findings demonstrate the effect of spatio-temporal patterns of subthreshold synaptic potentials on cortical processing and the shaping of perception.

Almost a century ago, Wertheimer¹ and Kenkel² made the striking observation that non-moving stimuli can give rise to motion perception⁹. Assuming that the facilitatory effect of the cue spread gradually in space and time^{10–12}, the sensation of motion evoked by a subsequently flashed bar during the line-motion illusion³ could result from sequential suprathreshold activation, thus mimicking real motion drawing away from the local cue^{3,13} (Fig. 1a, b; Supplementary movie SI 1, page 1). Whereas the findings of several psychophysical studies implicate top-down involvement of higher cortical areas^{14–16}, others argue for bottom-up (that is, stimulus induced) mechanisms occurring at early processing stages^{16–18}. In none of those studies, however, were the neurophysiological

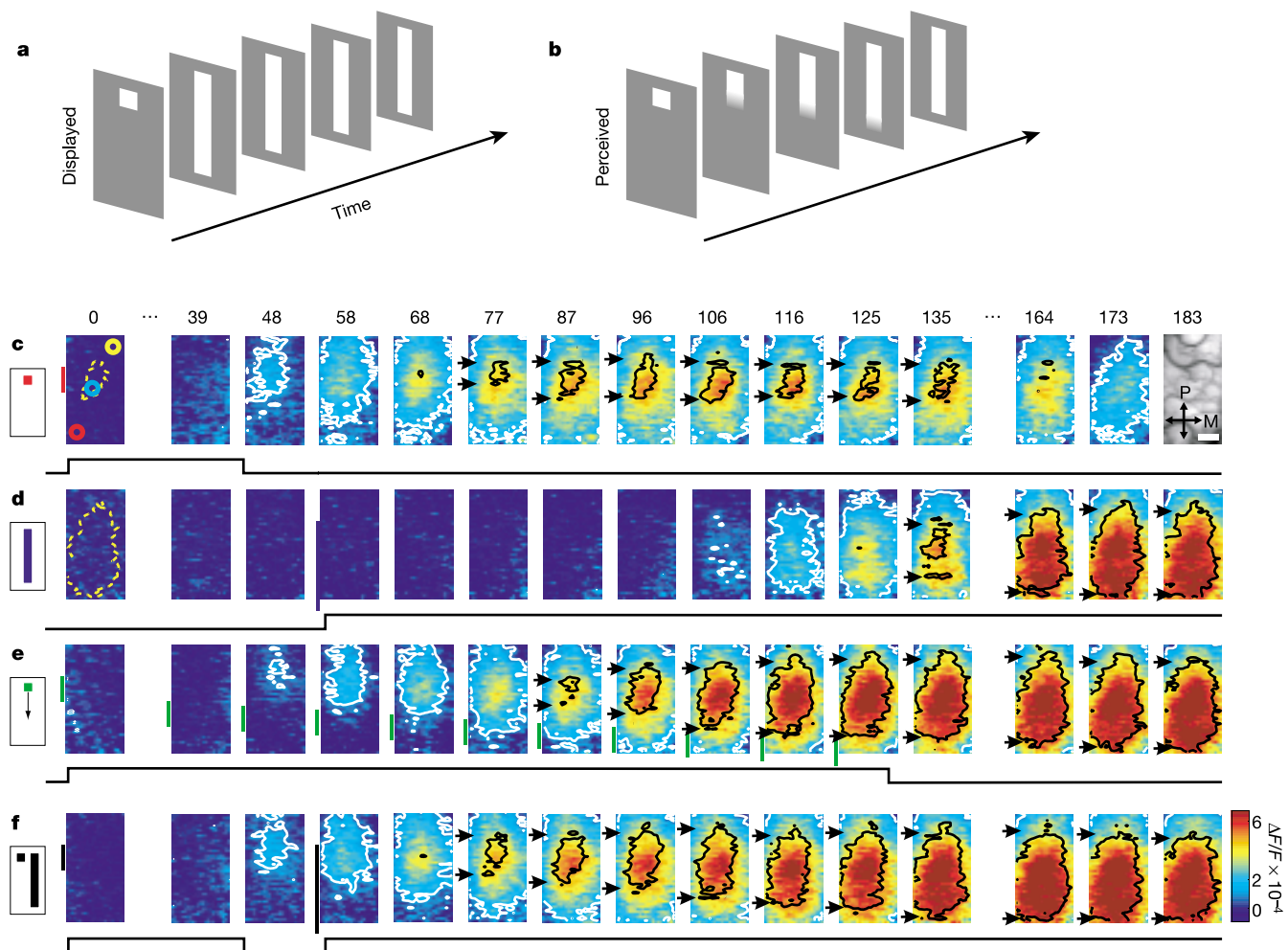


Figure 1 Cortical representations of stationary, moving, and illusory moving stimuli. The line-motion illusion. **a**, Square ('cue') presented before a bar stimulus. **b**, Subjects report illusory line-drawing. See Supplementary movie SI 1, page 1. **c–f**, Patterns of evoked cortical activity as a function of time. Yellow dotted contours approximate retinotopic representation of the stimuli; coloured circles indicate extracellular recording sites. White contours delimit low-amplitude activity (significance level; $P < 0.05$). The cortical area

imaged is shown at upper right. Scale bar, 1 mm; P, posterior; M, medial. Green vertical lines in **e** indicate estimated position of the stimuli along posterior–anterior axis. Time in milliseconds after stimulus onset is shown at the top. Stimulation time is shown at the bottom of each row. Colour scale indicates averaged fractional changes in fluorescence intensity ($\Delta F/F$). Stimuli: **c**, flashed small square. **d**, Flashed bar. **e**, Moving small square (32° s^{-1}). **f**, Line-motion paradigm. 22 repetitions were averaged.

mechanisms creating such illusory motion examined directly.

To identify neuronal mechanisms that produce illusory motion, we employed voltage-sensitive dye (VSD) imaging⁴ that allows sensitive monitoring of the changes in membrane potential, and particularly sub- and suprathreshold synaptic potentials^{4,19–21}, with high spatial and temporal resolutions⁶. Figure 1c depicts the cortical VSD response evoked by the cue alone (small square). As the VSD signal is sensitive primarily to synaptic potentials rather than to spikes^{19–21}, we searched for a way to delineate the spiking regions in the resulting optical maps. From basic principles of the retinotopic (spiking) organization of area 18, and because of the known spread of subthreshold synaptic potentials far beyond that area^{4,5,21}, it was obvious that the spiking region should be located in the area of high VSD amplitudes. Therefore, the dynamic behaviour of evoked low- and high-amplitude activity was analysed separately. Low-amplitude activity spread far beyond the retinotopic representation of the square, at 0.09 m s^{-1} ($n = 19$), consistent with conduction velocity along horizontal connections^{5,22}. In contrast, high-amplitude activity (encircled by black contours) showed negligible lateral extension (Fig. 1c arrows), after an initial $\sim 20 \text{ ms}$ period of spread. Single-unit recordings (coloured circles; Fig. 1c, first frame) showed that spiking activity evoked by the square was limited to the high-amplitude area (Fig. 2a, b, discussed below). The second stationary stimulus, the bar flashed alone 60 ms later, yielded a similar finding (Fig. 1d). High-level activity delineated a circumscribed region representing the elongated shape of the bar almost at once (arrows). In addition, activity within this cortical region consistently had the shortest latencies (34–37 ms; Supplementary Fig. S2).

To examine how actual motion is represented across the imaged cortical area, the small square was moved across exactly the same region as that covered by the bar (Fig. 1e). We again observed a rapid

spread of low-amplitude activity. However, in contrast to the stationary conditions, the moving square evoked high-amplitude activity propagating anteriorly, reflecting the motion trajectory of the square (lower arrows; Supplementary Fig. S1 3 for squares moving at different speeds and electrophysiological recordings). In summary, these findings suggest that regions of high optical activity correspond to the spiking representations of the stimuli. We therefore refer to those cortical regions as ‘spike discharge’ (SD) zones.

Next, we investigated the spatio-temporal pattern of activity evoked by the Hikosaka³ line motion paradigm (Fig. 1f). In contrast to the conditions in which either the square or the bar was flashed alone, the SD zone did not remain stable but was gradually drawn out towards the end of the cortical bar representation (lower arrows). Thus, although these two stimuli were stationary, the anterior portion of the high-activity region propagated similarly to the moving stimulus (Fig. 1e), suggesting that cortical correlates of illusory line motion were directly visualized here (Supplementary movie S1 1, page 2; another example in Supplementary Fig. S1 4).

To verify this suggestion, we imaged the responses to stimuli that were drawn out to full bar length at a speed of 32° s^{-1} , thus mimicking the perceived illusory motion (Fig. 2). The spatio-temporal patterns evoked by this high-speed real motion closely resembled the activity induced by the line-motion stimuli (Fig. 2a, second and third rows). The link between such cortical spatio-temporal patterns and object physical speeds is shown in a subsequent experiment where the bar was drawn out more slowly (16° s^{-1} , fourth row Fig. 2a). In this case, the speed of propagating SD activity was slower than for line motion, demonstrating that the expanding SD zone carries information about motion

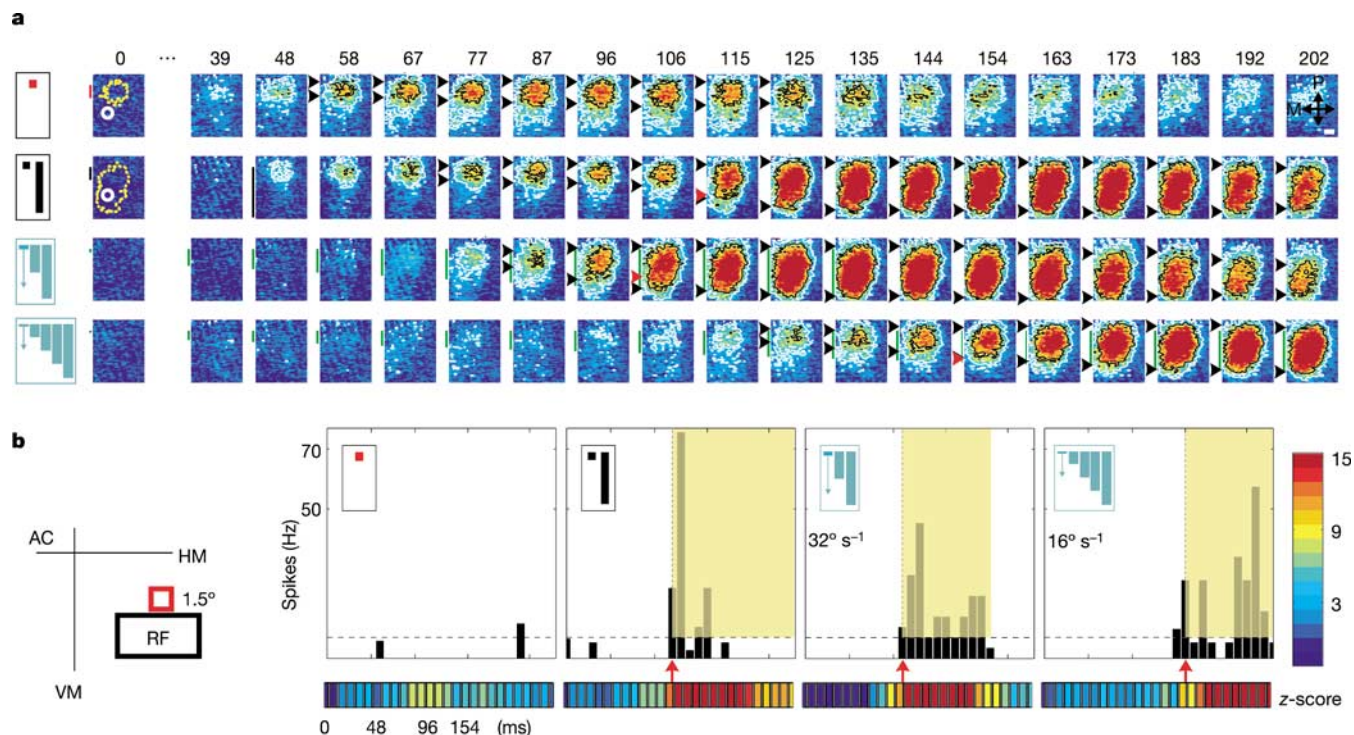


Figure 2 Both illusory and real motion evoke propagating spiking discharge (SD) zones. **a**, From top to bottom: VSD images patterns of activity evoked by flashed square, line-motion, and drawn-out bar at 32° s^{-1} and 16° s^{-1} . Multi-unit spiking activity was recorded simultaneously at the location marked by the white circle at frame 0; yellow dotted contours at frame 0 approximate sizes of the SD zone representing square and bar (left). **b**, The neurons' receptive field at the bottom left (RF, black), stimulus position and size (red); VM, HM, vertical, horizontal meridian; AC, area centralis; PSTHs of spiking

responses to the respective conditions are on the right. Dashed line depicts significance ($z\text{-score} = 2$; 32 repetitions). Spikes are detected as soon as activity reaches the SD threshold at the position of the electrode (red arrows). Bottom shows $z\text{-scores}$ of the VSD signal at electrode position. Area in yellow depicts the period in which the optical signal at the extracellular recording site is above SD level as shown by the three red triangles in the VSD patterns (**a**).

speed. Simultaneous VSD imaging and multi-unit recordings again confirmed that the SD zone corresponded to cortical regions containing propagating suprathreshold (spiking) activity whereas activity outside its border was subthreshold (Fig. 2b).

To specify the differences in the cortical representations of the various stimuli, we examined the time courses of activity in three contiguous regions of interest (ROIs; squares in insets of Fig. 3a). Activity evoked by the flashed square alone (red) was detected at all sites, but was smaller and slower at more anterior cortical positions. In sharp contrast, the flashed bar alone (blue) evoked activities exhibiting identical latencies and amplitudes within all three ROIs. Evidently, the two lower ROIs were outside the retinotopically activated area for the square and inside for the bar, containing activity that reached the SD level for many neurons (marked in yellow). The responses to the moving square, at the three ROIs, passed the SD level at successive times, thus representing the motion (green, Fig. 3b). Figure 3c shows the time course of activity evoked by the line-motion stimulus (black). As expected, at response onset the responses were similar to those evoked by the cue alone (compare black and red). However, during the time interval when the response to the stationary square alone turned into a plateau, activity for the line-motion stimuli continued to rise, crossing SD level at successive times, just as observed for motion. The latency of the response to the subsequently presented bar was strikingly shorter than the response to the bar alone (arrow in Fig. 3c, ~18 ms). Similar results were obtained in other experiments (23 ± 6 ms, mean $\pm$ s.d., $n = 4$), underscoring the impact of subthreshold activity on time of spiking evoked by a subsequent stimulus.

To further characterize the nature of the synaptic interactions between activities evoked by the cue stimulus and the bar response, we compared the time course of line-motion evoked activity with the calculated linear combination of its two components presented alone. The traces in Fig. 3d show the measured (black) and the calculated responses (pink), revealing two deviations from linearity of synaptic interactions. Activity was first increased by 17% (other

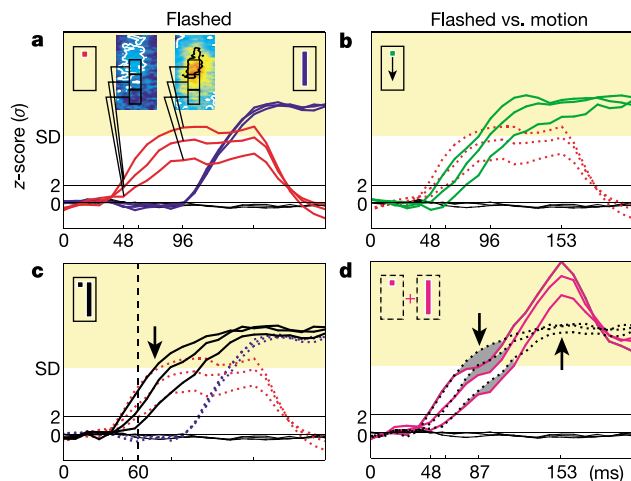


Figure 3 Time course of activity evoked by the stationary bar is largely affected by the preceding stationary cue. Responses from three contiguous cortical regions of interest (ROIs, black rectangles in insets at top). Yellow area marks activity above SD level; significance level (z -score = 2) is shown by a horizontal black line. **a**, Small square (red), bar (blue). **b**, Moving square (green; 32° s^{-1}), flashed square (red). **c**, Flashed square (red), flashed bar (blue), and line-motion condition (black). Arrow points to short latency of the response to the subsequently flashed bar; vertical dashed line shows time of presentation. **d**, Linear superposition of responses evoked by square and bar alone (pink) compared with the line-motion condition (black). Grey areas outline early deviation from a linear response. Thin black traces in **a–d** show time courses of blank conditions, indicating the excellent signal to noise ratio obtained.

experiments: $24 \pm 10\%$, mean $\pm$ s.d.; $n = 4$), and ~ 50 ms later, activity was decreased by as much as 32% ($31 \pm 4\%$; $n = 4$). An increase in the VSD signal could either reflect increase in excitation or withdrawal of inhibition. The net effect appeared 'facilitatory' in the early response phase whereas overall 'suppression' dominated at later times^{12,23,24}.

Do the characteristics of cortical subthreshold spread account for the illusory motion? Level of activation and speed of propagation of subthreshold activity co-varied: the higher the activation, the slower the propagation speed (Fig. 4a). This could arise from a gradual increase in inhibition²³, or from a gradual decrease in the synaptic density/efficacy of horizontal projections at further intracortical distances^{25,26}. The velocity of the subthreshold spread evoked by the cue was found to be similar to the velocity of SD activity produced when the line-motion stimuli are presented (Fig. 4b). It is therefore not a coincidence that propagating activity representing illusory motion was similar to SD activity evoked by real bar drawing or a moving square at a very specific speed (Fig. 4c). In contrast, no SD propagation was produced by a stationary bar or square alone (Fig. 4d; see Supplementary Information SI 5). Changing cue characteristics altered spatio-temporal patterns in a way consistent with psychophysical studies: first, when flashing a square with lower luminance the threshold for SD activity was reached later (Fig. 4e; Supplementary Fig. SI 6), in line with a weaker percept of illusory line motion¹³. Second, motion was not observed when the delay between the preceding square and the following bar exceeded 350 ms (Fig. 4f). Third, no line-motion correlates were found for spatial distances between square and bar beyond 6° in the visual field (not shown). Altogether, these findings confirmed the spatio-temporal limits of low-level processes participating in illusory line motion as demonstrated psychophysically^{16–18}. We therefore suggest that characteristics of subthreshold spread in early visual

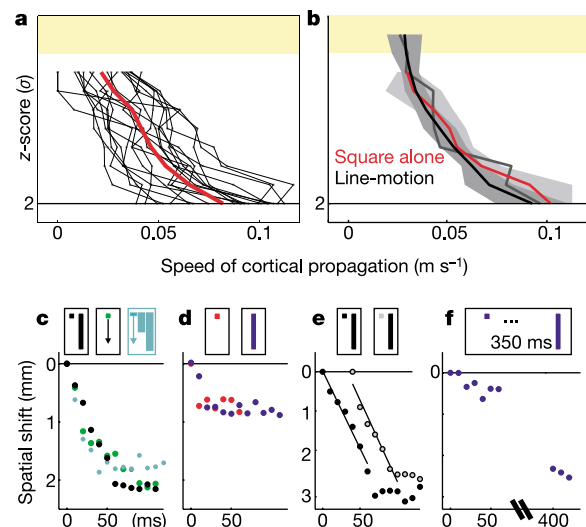


Figure 4 SD activity propagates only when real or illusory motion occurs.

a, b, Deceleration of propagation speed with increasing amplitudes (see Methods). **a**, Speeds of propagation for flashed squares (17 hemispheres), red trace shows average. Evidently, for the stationary square, no speed exists above SD activity level (yellow). **b**, Line motion (black) compared to the square alone (red; across four experiments). A square of lower luminance (see Methods) does not change propagation speed (grey line, 33 repetitions). Above SD level, responses to the line motion paradigm continue propagating at similar speed. **c–f**, Propagation of the SD zone (see Methods). **c**, Propagation of SD activity is evident for line motion (black), the moving square (green), and bar drawing (cyan). **d**, After an initial small expansion, no shift is observed for the flashed-stationary square and bar alone (red, blue). **e**, Reducing the square's contrast (grey) causes a lagged start of line motion. **f**, An interval of 350 ms between square and bar shows no line-motion effect; no SD activity is detectable at 70–400 ms.

areas shown here are cortical correlates of the bottom-up mechanism underlying the line-motion illusion, without need of attention^{18,27–29}.

It has been reported that when multiple inducers (or different stimulus features like colour, contrast, or texture) are used there is evidence for a second, top-down component that operates on a slower timescale^{16,18,30}. Furthermore, the illusory line motion can be modulated by attention²⁹ or can be voluntarily induced by non-retinotopic mechanisms¹⁶ even across sensory modalities¹⁵ (but see ref. 29, and Supplementary Information SI 7). Such controversies in psychophysical studies underscore the importance of physiological measurements. These top-down processing and attentional mechanisms^{14–16,29} remain to be explored in behaving subjects, by bringing together psychophysics and fast functional imaging. As onset of any stimulus will create an intricate spatio-temporal pattern of spreading subthreshold synaptic activity, we conclude that subthreshold synaptic spread must serve a fundamental processing strategy that strongly affects the spatio-temporal integration of any natural sensory input. □

Methods

Animals

All surgical and experimental procedures were in accordance with NIH guidelines. Details are described elsewhere⁶. Briefly, recordings were performed in 15 adult cats, initially anaesthetized with ketamine (15 mg kg⁻¹) and xylazine (1 mg kg⁻¹), supplemented with atropine (0.05 mg kg⁻¹). After tracheotomy, cats were artificially respired and were anaesthetized with 1–1.5% halothane (0.8–1% during recordings) in a 1:1 mixture of O₂ and N₂O. The skull was opened above area 18 and the dura was resected. Paralysis was induced with pancuronium bromide (0.2 mg kg⁻¹ h⁻¹, intravenously). The eyes were covered with zero-power smooth contact lenses. External lenses were used to focus the eyes on the screen. Eyes were converged using a prism in front of the eye ipsilateral to the recorded hemisphere. To control for eye drift, the position of the area centralis was repeatedly measured. Eye movements were completely eliminated under these experimental conditions.

Stimulation, imaging and electrical recordings

The cortex was stained for 2.5–3 h using the oxonol dye RH-1691. A FUJIX HR Deltaron 1700 camera was used for data acquisition. Combined intracellular recordings and VSD imaging *in vivo* have demonstrated that the VSD signal reports amplitude, direction, and propagation speed of synaptic potentials in layer 2/3 of cortical neurons^{20,21} (in a linear fashion, without saturation). Details of the method have been previously described⁶. Stimuli were displayed binocularly with the VSG series 3 (Cambridge Research Systems) on a monitor located 57 cm from the cat's eyes, at a refresh rate of 150 Hz. The achromatic luminance of the stimuli was 52.5 cd m⁻² (26.2 cd m⁻² when testing for lower luminance). Stimuli were presented on a grey background (10.5 cd m⁻²). Small squares of light were flashed for 50 ms, and their locations on the screen were determined according to single-unit recordings. Flash duration of the vertical flashed bar was 130 ms. Varying the cue lead times between 60 and 100 ms did not significantly alter propagation speed, in accordance with psychophysical studies pointing to a stable illusory line motion within the same range of bar delays^{3,17,18}. Hikosaka *et al.*³ report the strongest illusion when presenting an inducing cue 100–200 ms before flashing a line. Von Grünau *et al.*¹⁸ discriminate two contributions to illusory line-motion: a pre-attentive fast-acting process^{16,30} followed by a slower, probably top-down process, showing strongest motion effects for bars presented 200–300 ms after the cue. Therefore, to test the limits of line-motion induction, we increased the bar delay up to 350 ms and presented the bar separated by 3–6° from the square. Stimuli were located between 4° and 5.5° eccentricity in different experiments and were presented pseudo-randomly. Single-unit and multi-unit recordings were performed as described elsewhere^{7,8}. A recording time of about 1 min surrounded each stimulus presentation.

Data analysis

To remove slow common noise, we obtained the time courses of evoked activity by subtracting the average value at each pixel before stimulus onset from each pixel value. To correct for the fast heart beat noise and respiration noise and to obtain the activation maps (in units of fractional change of fluorescence intensity $\Delta F/F$), we divided the resulting values by the activity recorded during the 'no-stimulus' condition⁶. Such data reflect evoked population neuronal activity, as confirmed by intracellular recordings^{19,20}.

As shown by the black traces in Fig. 3a–d, the signal to noise ratio in these VSD experiments is excellent. To calculate statistical significance (*z*-score), we divided each pixel by its standard deviation calculated over the blank conditions, after subtracting its mean. No spatio-temporal filters were used. We computed the speeds of cortical activity within space–time diagrams across the entire time courses of the evoked responses. Activity was averaged for each single time frame across the *x*-dimension, providing us with a single space dimension as a function of time (space–time diagram). Contours were calculated at 10 different *z*-score levels of activity (scaled in steps of 10% from a *z*-score of 2 to the maximum *z*-score observed) on such diagrams. The slope of each contour (computed by linear regression) within such a space–time diagram corresponds to the speed of propagating activity at a given *z*-score.

Received 26 November 2003; accepted 5 February 2004; doi:10.1038/nature02396.

- Wertheimer, M. Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.* **61**, 161–265 (1912).
- Kenkel, F. Untersuchungen über den Zusammenhang zwischen Erscheinungsgröße und Erscheinungsbewegung bei einigen sogenannten optischen Täuschungen. *Z. Psychol.* **67**, 358–449 (1913).
- Hikosaka, O., Miyauchi, S. & Shimojo, S. Focal visual attention produces illusory temporal order and motion sensation. *Vision Res.* **33**, 1219–1240 (1993).
- Grinvald, A., Angiller, L., Freeman, J. A., Hildesheim, R. & Manker, A. Real-time optical imaging of naturally evoked electrical activity in intact frog brain. *Nature* **308**, 848–850 (1984).
- Grinvald, A., Lieke, E., Frostig, R. & Hildesheim, R. Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J. Neurosci.* **14**, 2545–2568 (1994).
- Shoham, D. *et al.* Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes. *Neuron* **24**, 791–802 (1999).
- Albus, K. A quantitative study of the projection area of the central and the paracentral visual field in area 17 of the cat. I. The precision of the topography. *Exp. Brain Res.* **24**, 159–179 (1975).
- Tusa, R. J., Rosenquist, A. C. & Palmer, L. A. Retinotopic organization of areas 18 and 19 in the cat. *J. Comp. Neurol.* **185**, 657–678 (1979).
- Kanizsa, G. Sulla polarizzazione del movimento gamma. *Arch. Psicol. Neurol. Psichiatr.* **3**, 224–267 (1951).
- Posner, M. I., Snyder, C. R. R. & Davidson, B. J. Attention and the detection of signals. *J. Exp. Psychol.* **109**, 160–174 (1980).
- Sagi, D. & Julesz, B. Enhanced detection in the aperture of focal attention during simple discrimination tasks. *Nature* **321**, 693–695 (1986).
- Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T. & Norcia, A. M. Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* **391**, 580–584 (1998).
- von Grünau, M., Racette, L. & Kwas, M. Measuring the attentional speed-up in the motion induction effect. *Vision Res.* **36**, 2433–2446 (1996).
- von Grünau, M. & Faubert, J. Intraattribute and interattribute motion induction. *Perception* **23**, 913–928 (1994).
- Shimojo, S., Miyauchi, S. & Hikosaka, O. Visual motion sensation yielded by non-visually driven attention. *Vision Res.* **37**, 1575–1580 (1997).
- Hikosaka, O. & Miyauchi, S. Voluntary and stimulus induced attention detected as motion sensation. *Perception* **22**, 517–526 (1993).
- Steinman, B. A., Steinman, S. B. & Lehmkuhle, S. Visual attention mechanisms show a center-surround organization. *Vision Res.* **35**, 1859–1869 (1995).
- von Grünau, M., Dube, S. & Kwas, M. Two contributions of motion induction: a preattentive effect and facilitation due to attentional capture. *Vision Res.* **36**, 2447–2457 (1996).
- Sterkin, A. *et al.* Real-time optical imaging in cat visual cortex exhibits high similarity to intracellular activity. *Neurosci. Lett.* **51**, S41–S41 (1998).
- Grinvald, A. *et al.* in *Modern Techniques in Neuroscience Research* (eds Windhorst, U. & Johansson, H.) 893–969 (Springer, New York, 1999).
- Petersen, C., Grinvald, A. & Sakmann, B. Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured *in vivo* by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. *J. Neurosci.* **23**, 1298–1309 (2003).
- Bringuier, V., Chavane, F., Glaeser, L. & Frégnac, Y. Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* **283**, 695–699 (1999).
- Hirsch, J. A. & Gilbert, C. D. Synaptic physiology of horizontal connections in the cat's visual cortex. *J. Neurosci.* **11**, 1800–1809 (1991).
- Jancke, D. *et al.* Parametric representation of retinal location: Neural population dynamics and interaction in cat visual cortex. *J. Neurosci.* **19**, 9016–9028 (1999).
- Kisvárdy, Z. F., Toth, E., Rausch, M. & Eysel, U. T. Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cereb. Cortex* **7**, 605–618 (1997).
- Seriès, P., Georges, S., Lorenceau, J. & Frégnac, Y. Orientation dependent modulation of apparent speed: a model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. *Vision Res.* **42**, 2781–2797 (2002).
- Kawahara, J., Yokosawa, K., Nishida, S. & Sato, T. Illusory line motion in visual search: Attentional facilitation or apparent motion? *Perception* **25**, 901–920 (1996).
- Cavanagh, P., Arguin, M. & von Grünau, M. Interattribute apparent motion. *Vision Res.* **29**, 1197–1204 (1989).
- Downing, P. E. & Treisman, A. M. The line-motion illusion: Attention or impletion? *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 768–779 (1997).
- Nakayama, K. & Mackeben, M. Sustained and transient components of focal visual attention. *Vision Res.* **29**, 1631–1647 (1989).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Hildesheim for synthesizing the dye, RH-1692; C. Wijnbergen, D. Ettner and Y. Toledo for technical assistance; I. Lampl for programming the spike recording software; and A. Arieli, A. Gupta, C. E. Schreiner, D. Sharon, D. Sagi, E. Seidemann, H. Slovlin, I. Vanzetta, S. Ullman and Y. Frégnac for comments and discussions. This work was supported by the Grodzky Center, the Goldsmith, Korber and ISF Foundations and a BMBF/MOS and NIH grants (to A.G.). D.J. was supported by the Minerva Foundation, Germany, and E.C. by a Marie Curie EU fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J. (jancke@neurobiologie.ruhr-uni-bochum.de).

A resource for large-scale RNA-interference-based screens in mammals

Patrick J. Paddison^{1*}, Jose M. Silva^{1*}, Douglas S. Conklin^{1*†}, Mike Schlachach^{2†}, Mamie Li², Shola Aruleba¹, Vivekanand Balija¹, Andy O'Shaughnessy¹, Lidia Gnoj¹, Kim Scobie¹, Kenneth Chang¹, Thomas Westbrook^{2†}, Michele Cleary³, Ravi Sachidanandam¹, W. Richard McCombie¹, Stephen J. Elledge^{2†} & Gregory J. Hannon¹

¹Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

²Department of Biochemistry, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

³Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA

* These authors contributed equally to this work

† Present addresses: Department of Biomedical Sciences, Center for Functional Genomics, University at Albany, East Campus, B342A, One University Place, Rensselaer, New York 12144-2345, USA (D.S.C.); Department of Genetics, Harvard Partners Center for Genetics and Genomics, Harvard Medical School Room 158D, NRB, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA (M.S., T.W. and S.J.E.)

Gene silencing by RNA interference (RNAi) in mammalian cells using small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs) has become a valuable genetic tool^{1–10}. Here, we report the construction and application of a shRNA expression library targeting 9,610 human and 5,563 mouse genes. This library is presently composed of about 28,000 sequence-verified shRNA expression cassettes contained within multi-functional vectors, which permit shRNA cassettes to be packaged in retroviruses, tracked in mixed cell populations by means of DNA 'bar codes', and shuttled to customized vectors by bacterial mating. In order to validate the library, we used a genetic screen designed to report defects in human proteasome function. Our results suggest that our large-scale RNAi library can be used in specific, genetic applications in mammals, and will become a valuable resource for gene analysis and discovery.

In invertebrates, RNAi has been harnessed as a powerful genetic

tool to reduce endogenous gene expression through programming organisms or cells with homologous double-stranded (ds)RNA (reviewed in ref. 1). In *Caenorhabditis elegans*, this approach has been used in large-scale, genome-wide screens^{2–5}. With the advent of RNAi in mammals and the refinement of techniques to trigger gene silencing, expression from any human or mouse transcript can, in principle, be inhibited using siRNAs or shRNAs.

To facilitate the use of RNAi as a genetic tool in mammals, we have constructed a large-scale library of RNAi-inducing shRNA expression vectors targeting human and mouse genes. We began by testing a number of variables that might affect shRNA performance. We previously showed that shRNAs containing 29 nucleotides of dsRNA and simple loop structures are effective silencing triggers when expressed from U6 small nuclear RNA promoters^{6,7}. Other published accounts of shRNA-mediated inhibition include differences in promoter choice (reviewed in ref. 8), orientation of sense and antisense strands, length of RNA duplex (19–29 nucleotides), loop structure, and the addition of a 27-nucleotide U6 leader sequence⁹. After comparing each of these variables with a series of eight shRNAs targeting firefly luciferase or green fluorescent protein (GFP), we found that RNA polymerase III promoters were largely interchangeable (Supplementary Fig. 1a, b), that 29-nucleotide hairpins were more effective than shorter shRNAs (Supplementary Fig. 1a and data not shown), that changes in loop structure had minimal effects (data not shown), and that the addition of the U6 leader sequence had a positive effect (Supplementary Fig. 1c). The finalized hairpin design for this library is presented in Supplementary Fig. 2a. It contains a 27-nucleotide U6 leader sequence, followed by 29 base pairs (bp) of dsRNA and a 4-nucleotide loop. Recent studies have suggested that each shRNA is processed from its stem end by Dicer through a single cleavage event (D. Siolas and G.J.H., unpublished data). Thus, the combination of Drosha and Dicer processing is predicted to generate a precisely defined siRNA from each 29-nucleotide shRNA expression vector.

shRNAs were designed covering approximately 10,000 human and 5,000 mouse genes with between three and nine constructs each (Supplementary Fig. 2b). Only coding sequences were targeted, and each shRNA was chosen such that it contained >3 mismatches to any other gene. Where possible, shRNAs had sequence identity to the mouse orthologue of the targeted gene. In pilot experiments,

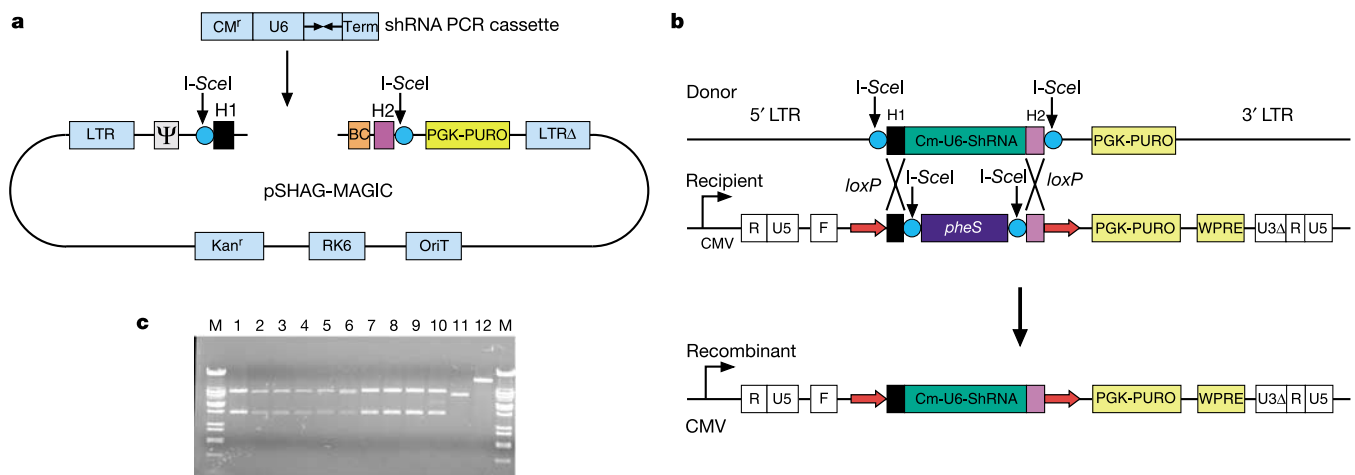


Figure 1 pSHAG-MAGIC shRNA cassette movement strategy. **a**, Map of the pSHAG-MAGIC vector. CM^r, chloramphenicol-resistance gene; Kan^r, kanamycin-resistance gene; LTR, long terminal repeat; OriT, origin of transfer. **b**, A diagrammatic representation of DNA exchanges occurring once the pSHAG-MAGIC donor vector has been transferred to cells containing a recipient vector by mating. In this case pLenti-LoxP (M.L. and S.J.E., unpublished data) is the recipient vector. WPRE, woodchuck hepatitis B post-transcriptional responsive element. **c**, Plasmids from ten independent colonies (post-

mating) were digested with *Nde*I and the digestion products were separated on a 0.5% agarose gel (lanes 1–10). The parental plasmids, pSHAG-MAGIC (lane 11) and pLenti-LoxP (lane 12), each contain a single *Nde*I site, and on cleavage generate a 7.0- or 10.6-kilobase (kb) fragment, respectively. The pSHAG-MAGIC vector contains a unique *Nde*I site in the U6 promoter, which is transferred into the recipient vector. The correct mating product generates two fragments of 3.5 kb and 7.4 kb. M, marker.

between 25–75% of cloned shRNAs contained significant mutations, which arose during chemical synthesis. As a result we sequence-verified each shRNA in the library (see Supplementary Fig. 2b).

Our current sequence-verified library has a representation of 9,610 human and 5,563 mouse genes. This corresponds to 28,659 shRNAs targeting human genes and 9,119 shRNAs targeting mouse genes. Examples of library coverage for selected functional categories of human genes are presented in Supplementary Table 1. The most thoroughly represented functional groups in our current library are kinases and phosphatases, in which approximately 85% of all known kinases and phosphatases are targeted by three or more hairpins. Most other functional classes contain between 30–60% coverage with three or more hairpins, and >80% of genes in functional categories listed in Supplementary Table 1 are targeted by at least one sequence-verified hairpin.

The shRNA expression library has been constructed in a vector that contains a number of convenient design features (pSHAG-MAGIC, Fig. 1a). The vector is capable of producing self-inactivating murine-stem-cell-virus (MSCV) particles in commonly available retroviral packaging lines. We have also incorporated a new *in vivo* subcloning technology. This method is called 'Mating-Assisted Genetically Integrated Cloning' (MAGIC) (M.L. and S.J.E., manuscript in preparation). The MAGIC system consists of a donor vector (the library vector), in which the fragment of

interest is flanked by two different 50-bp homology regions, H1 and H2, which in turn are flanked with linked I-SceI sites. The donor vector also includes an F' origin and a conditional origin of replication (RK6γ; Fig. 1a, b). The recipient vector, which also contains I-SceI-linked H1 and H2 sites surrounding a negative selectable marker (*pheS*), resides in a bacterial strain that contains an inducible I-SceI gene (Fig. 1b). After transfer of the donor vector into the recipient host by bacterial mating, I-SceI cleaves both donor and recipient vectors, and these breaks are healed by homologous recombination via the H1 and H2 sequences. Selection against the unrecombined recipient containing *pheS* and I-SceI sites and for the capture of the appropriate insert (chloramphenicol resistance) gives essentially 100% recovery of the desired plasmid. To test MAGIC transfer of shRNA clones, we developed a lentivirus recipient vector based on the FUV vector¹⁰. Using a test shRNA clone in the mating, we observed a mating efficiency of $>6 \times 10^6$ clones per ml of mixed bacteria. Restriction analysis showed that 10 out of 10 recombinants have the expected structure (Fig. 1c).

The shRNA library was designed to function for both genetic selections and screens. For selections conferring a growth advantage, the library can be used in a pooled fashion. Genetic screens (for example, for lethal events) can be carried out in a 96-well format; however, this method is labour intensive and time consuming. To facilitate such screens, we have adopted a DNA bar-coding strategy

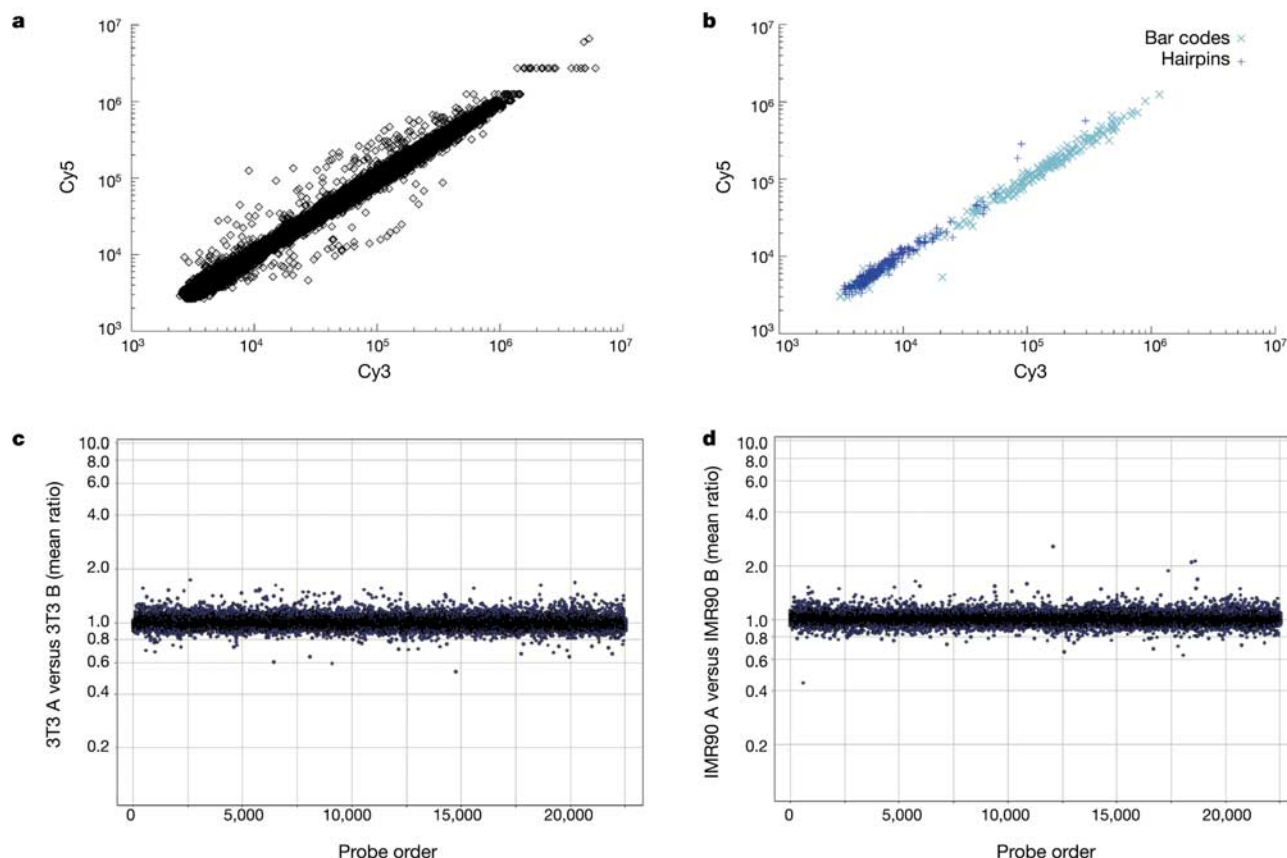


Figure 2 Microarray analysis of pSHAG-MAGIC library bar codes. **a**, Self-self hybridization of pSHAG-MAGIC library bar codes obtained from approximately 15,000 library plasmids prepared from *E. coli*. The DNA microarray (Agilent Technologies) is composed of 20,241 complementary 60-nucleotide oligonucleotides out of a total of 22,575 elements on the array, including controls; the remaining represent various positive and negative controls. Cy-labelled cRNA for library bar codes was generated and competitively hybridized as described in the Supplementary Methods. **b**, An analysis of a subset of 255 60-nucleotide bar codes from **a** versus 255 60-nucleotide shRNA probes. Each 60-nucleotide shRNA probe contains a direct repeat of the 29-nucleotide gene-targeting sequence with a 2-nucleotide spacer. **c**, Microarray analysis of pSHAG-MAGIC

bar codes from transduced NIH3 T3 cells. Two pools of 2×10^7 cells (A and B) were infected separately and harvested 48 h after infection. Bar codes were processed and competitively hybridized as described in the Methods. The x axis is an index of all the elements (22,575), bar codes, hairpins and controls on the array. The y axis shows the mean measured ratios from two experiments (forward and reverse colour orientation) of 3T3 pool A hybridized against 3T3 pool B, plotted on a log scale with the numbers indicating fold change. Most probes, with the exception of a few, scatter around 1.0 (no change) with a range between 0.5 (twofold decrease) and 2.0 (twofold increase). **d**, Microarray analysis of pSHAG-MAGIC bar codes in transduced normal human diploid fibroblasts (IMR90) cells carried out as in **c**.

that has been used previously in *Saccharomyces cerevisiae* deletion collections, to follow individual mutants in complex populations via microarray analysis^{11–13}. We have linked a unique 60-nucleotide DNA bar code to each shRNA vector to allow us to follow the fate of shRNAs in populations of virally transduced cells. To monitor relative frequencies of individual shRNA species, we designed microarrays (Agilent Technologies) containing 20,241 60-nucleotide DNAs complementary to each shRNA bar code.

Sequence analysis of library clones revealed that a fraction (about 4%) contained orphan shRNAs without bar codes. We therefore sought to determine whether we could use the shRNA sequence as a bar code. We chose 255 non-orphan shRNAs and deposited on arrays both associated bar codes and oligonucleotides designed to have 60 hybridize to the shRNA itself. Such oligonucleotides were

designed nucleotides in order to contain direct repeats of the 29-nucleotide antisense sequence. Cy-labelled antisense RNA was generated from about 15,000 library constructs such that it contained the reverse complement of the bar code and the entire shRNA. A linear log-log plot of a self hybridization revealed consistent cRNA production and hybridization from the plasmid library (Fig. 2a). However, hairpins gave substantially lower hybridization signals, essentially at background levels, relative to the corresponding bar codes (Fig. 2b), indicating that using hairpins as probes is not an optimal bar-coding strategy under these conditions.

We next asked whether bar codes could be used to report the representation of individual shRNAs in transduced mammalian cells. Normal human diploid fibroblasts (IMR90) or NIH3T3 cells

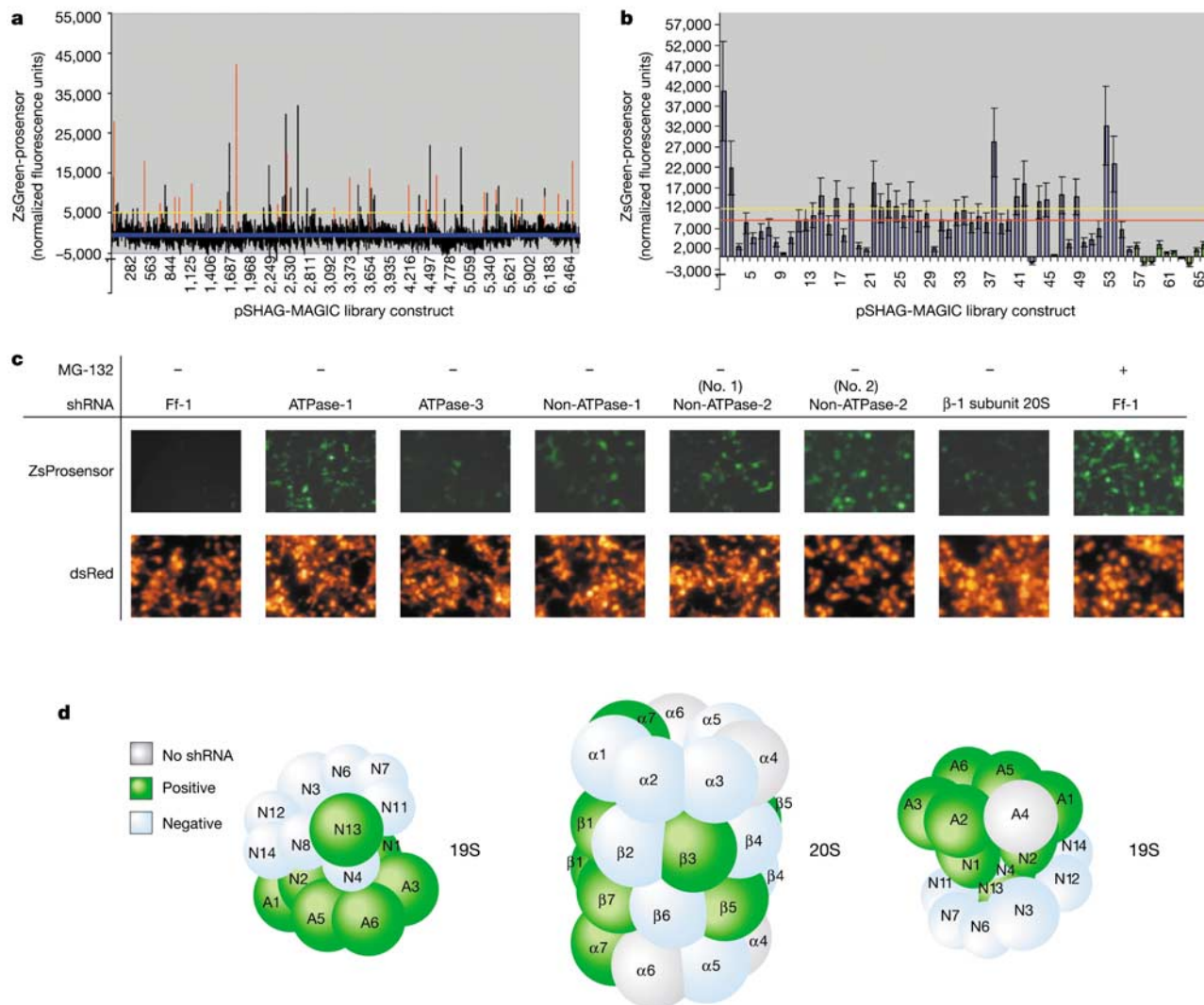


Figure 3 A reverse genetic screen for defects in human proteasome function. **a**, A graph of relative ZsGreen fluorescence for all pSHAG-MAGIC clones transfected. Red vertical bars indicate 22 positively scoring proteasome shRNAs corresponding to 15 known proteasome subunits, whereas black bars indicate library shRNAs. The yellow horizontal bar indicates a cut-off that was set based on control experiments. **b**, A replicate transfection experiment as in **a** using a set of 22 positively scoring proteasome hairpins and 33 that failed to score (blue bars), along with non-proteasome hairpin controls (green bars). These experiments were carried out in triplicate. All 22 of the proteasome hairpins with a positive score from the original screen were re-tested and also achieved a positive score. None of the 33 non-scoring proteasome hairpins from the original screen managed to match the score of the 22 positively scoring proteasome shRNAs from the first round

(yellow line). However, in this experiment 36 of 55 scored well above (~2 s.d.) the mean background (red line). **c**, Fluorescence microscopy images showing representative results for individual proteasome hairpins as carried out in **a** and **b**. A shRNA targeting firefly luciferase (Ff-1) and small molecule proteasome inhibitor, MG-132 (Sigma), were used as negative and positive controls, respectively. **d**, A diagrammatic representation of the 26S proteasome colour-coded according to pSHAG-MAGIC library hits. Subunits coloured green had strong positive shRNA hits from the library in the primary screen. Grey subunits were not represented by any shRNAs in the approximately 7,000 tested. Subunits coloured blue were represented by at least 1 shRNA but did not score in the screen. For nomenclature see Supplementary Table 2.

were infected with retroviruses derived from the library population that was used for control hybridizations. To test for experimental variability, two populations of 2×10^7 cells (denoted A and B) were infected independently at a multiplicity of infection of approximately 1. Examination of colour-reversal experiments and plots comparing the relative intensity of the bar-code signals in these populations (Figs 2c, d) indicated that all steps of the procedure were highly reproducible, with the ratio of intensities varying by less than twofold in all but a few isolated cases for each cell line.

Finally, we sought to test the performance of the shRNA library in a biological context. We focused on an assay for which we could predict the recovery of a substantial number of shRNAs targeting genes in a known biological pathway. The 26S proteasome is the major non-lysosomal protease in eukaryotic cells. To search the library for shRNAs that compromise proteasome function, we used a reporter assay in which a fluorescent protein is coupled to a well-characterized degradation signal. The mouse ornithine decarboxylase (MODC) gene contains a PEST sequence that directs proteasomal degradation without the need for ubiquitination¹⁴. Roughly 7,000 shRNA expression plasmids, corresponding to approximately one-quarter of the complete library, were individually co-transfected with two expression constructs. The first encoded a *Zoanthus* green fluorescent (ZsGreen)-MODC degron fusion. In cases in which shRNAs compromised proteasome function, ZsGreen-MODC was expected to accumulate, giving a detectable signal. The second plasmid encoded *Discosoma* red fluorescent protein (DsRed). This permitted normalization of signals, which controlled for transfection efficiency.

An analysis of 6,712 shRNAs targeting 4,873 genes revealed approximately 100 RNAi constructs that increased the accumulation of ZsGreen-MODC (Fig. 3a). Twenty-two of these corresponded to 15 known proteasome subunits. As a secondary test,

these 22 putative proteasome-positive shRNAs, and an additional 33 shRNAs that targeted proteasome subunits but which had not scored in the original screen, were selected from the population. These were assayed in replicate transfections for the ability to increase the stability of the unstable fluorescent protein. Again, the 22 shRNAs scored positively, whereas the 33 proteasome shRNAs not detected in the initial screen scored less well. However, an additional 14 shRNAs did score above background in the focused assay (Fig. 3b, c).

Notably, many of the positive shRNAs targeted 19S base subunits, including five out of five ATPases and the two largest non-ATPases (1 and 2) (Fig. 3c, d; see also Supplementary Table 2). Compared with the 19S base, targeting subunits of the 19S lid or the 20S core produced a lower hit rate. However, two out of three of the most important catalytic components, located in subunits $\beta 1$ (peptidyl-glutamyl hydrolysing or caspase-like) and $\beta 5$ (chymotrypsin-like)^{15,16}, achieved a positive score in our assay (Fig. 3). The other two β -subunits that we identified are involved in important *cis* contacts between neighbouring subunits ($\beta 3$) and in *trans* contacts between the two β -rings ($\beta 7$)^{17,18}. In all cases tested, activation of the reporter was accompanied by a reduction in the expression of targeted proteasomal components (Fig. 4a and data not shown).

The c-Myc oncoprotein is a target for ubiquitin-mediated degradation by the proteasome¹⁹. As was observed for ZsGreen-MODC degron and bulk-ubiquitinated protein (not shown), c-Myc levels specifically increased upon transfection of cells with shRNAs that targeted proteasomal subunits (Fig. 4b).

Here, we report progress towards the construction of a genome-wide library of RNAi-inducing constructs for use in mammalian cells. At present the resource targets approximately 10,000 human and 5,000 murine genes, and it continues to expand. To examine the performance of the library, we have tested about one-quarter of its constituent clones (roughly 7,000 shRNA expression vectors) individually for the ability to inhibit the degradation of a direct proteasome target. Nearly 50% of the shRNAs in this collection that were expected to target proteasomal proteins were recovered as positives. The availability of this resource to the research community will open the door to the use of RNAi in mammalian systems as a large-scale tool for biological discovery. □

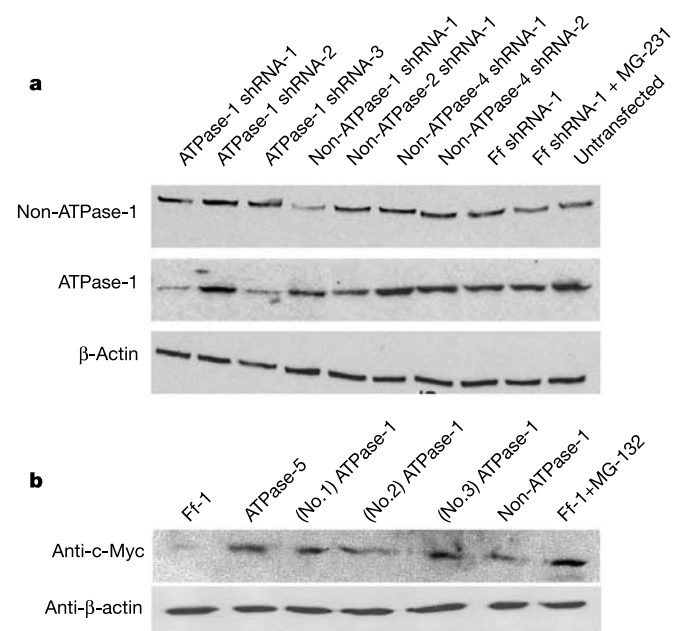


Figure 4 Further validation of selected pSHAG-MAGIC proteasome hairpins. **a**, Western blot showing specific suppression of the ATPase-1 and non-ATPase-1 of the 19S regulatory subunit in transiently transfected HEK293 cells. Cells were transfected with shRNAs as indicated. Knockdown of protein levels for shRNA-1 and -3 against ATPase-1 proteasome correlated with the severity of the relative scores in the pZsGreen-MODC accumulation assay. The lane labelled untransfected indicates the control lane where cells were not transfected. **b**, A western blot showing increased steady-state levels of endogenous c-Myc in HEK293 cells transiently transfected with library shRNAs as indicated. c-Myc is normally degraded by ubiquitin-mediated proteolysis in these cells¹⁹.

Methods

A detailed protocol for the design and creation of shRNA polymerase chain reaction (PCR) cassettes is available at <http://www.cshl.edu/public/SCIENCE/hannon.html>. Oligonucleotides were purchased from Sigma-Genosys. The pSHAG-MAGIC shRNA library was cloned in the MAGIC-competent *Escherichia coli* strain BW23474F⁺ DOT (M.L. and S.J.E., manuscript in preparation). The full sequence of the pSHAG-MAGIC vector and details of the mating protocol are available for download at the shRNA library website (<http://www.cshl.edu/public/SCIENCE/hannon.html>). Bar-code microarrays were synthesized by Agilent Technologies and were analysed as described in the Supplementary Methods.

Plasmids were prepared as described in the Supplementary Methods. Transfections were carried out in 96-well plates using plasmid mixtures described in Supplementary Methods. Fluorescence signals were read on a Victor2 plate reader. Signals in the green channel were normalized to transfection efficiency using customized scripts with fluorescence in the red channel serving as a normalization criterion. Cut-offs were assigned by using 16 independent control shRNA transfections to determine the range for a negative outcome.

Received 16 November 2003; accepted 26 January 2004; doi:10.1038/nature02370.

- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
- Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33**, 40–48 (2003).
- Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
- Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
- Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
- Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nature Genet.* **33**, 396–400 (2003).
- Paddison, P. J. & Hannon, G. J. siRNAs and shRNAs: skeleton keys to the human genome. *Curr. Opin. Mol. Ther.* **5**, 217–224 (2003).

9. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nature Biotechnol.* **20**, 505–508 (2002).
10. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
11. Birrell, G. W., Giaever, G., Chu, A. M., Davis, R. W. & Brown, J. M. A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *Proc. Natl Acad. Sci. USA* **98**, 12608–12613 (2001).
12. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
13. Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
14. Ghoda, L., Sidney, D., Macrae, M. & Coffino, P. Structural elements of ornithine decarboxylase required for intracellular degradation and polyamine-dependent regulation. *Mol. Cell. Biol.* **12**, 2178–2185 (1992).
15. Chen, P. & Hochstrasser, M. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **86**, 961–972 (1996).
16. Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U. & Wolf, D. H. The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **272**, 25200–25209 (1997).
17. Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. & Huber, R. The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 295–317 (1999).
18. Coux, O. An interaction map of proteasome subunits. *Biochem. Soc. Trans.* **31**, 465–469 (2003).
19. Kim, S. Y., Herbst, A., Tworowski, K. A., Salghetti, S. E. & Tansey, W. P. Skp2 regulates Myc protein stability and activity. *Mol. Cell* **11**, 1177–1188 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Moore and B. Simmons from Open Biosystems for their help in organizing and rearranging the library, and colleagues at CSHL and elsewhere (as indicated in Supplementary Table 1) as well as J. LaBaer and C. Perou for curating gene lists. G. Katari and J. Faith helped with bioinformatic analysis and shRNA choice, and members of the Lowe laboratory (CSHL) provided advice on vector optimization. This work was supported by an Innovator Award from the US Army Breast Cancer Research Program (G.J.H.), a contract from the National Cancer Institute (G.J.H.), grants from the NIH (G.J.H., W.R.M., S.J.E.) and the US Army Breast Cancer Research Program (G.J.H., D.S.C.), the Howard Hughes Medical Institute (S.J.E.), and by generous support from Oncogene Sciences and Merck. P.J.P. is an Arnold and Mabel Beckman Fellow of the Watson School of Biological Sciences and is supported by a predoctoral fellowship from the US Army Breast Cancer Research Program. J.M.S. is supported by a postdoctoral fellowship from the US Army Prostate Cancer Research Program. S.J.E. is an Investigator of the Howard Hughes Medical Institute. G.J.H. is a Rita Allen Foundation Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.org) or S.J.E. (selledge@genetics.med.harvard.edu).

A large-scale RNAi screen in human cells identifies new components of the p53 pathway

Katrien Berns^{1*}, E. Marielle Hijmans^{1*}, Jasper Mullenders¹, Thijn R. Brummelkamp¹, Arno Velds¹, Mike Heimerikx¹, Ron M. Kerkhoven¹, Mandy Madiredjo¹, Wouter Nijkamp¹, Britta Weigelt², Reuven Agami³, Wei Ge⁴, Guy Cavet⁴, Peter S. Linsley⁴, Roderick L. Beijersbergen¹ & René Bernards¹

¹Division of Molecular Carcinogenesis and Center for Biomedical Genetics, ²Division of Experimental Therapy, and ³Division of Tumor Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

⁴Rosetta Inpharmatics, Inc., 12040 115th Avenue NE, Kirkland, Washington 98034, USA

* These authors contributed equally to this work

RNA interference (RNAi) is a powerful new tool with which to perform loss-of-function genetic screens in lower organisms and can greatly facilitate the identification of components of cellular signalling pathways^{1–3}. In mammalian cells, such screens have been hampered by a lack of suitable tools that can be used on a large scale. We and others have recently developed expression

vectors to direct the synthesis of short hairpin RNAs (shRNAs) that act as short interfering RNA (siRNA)-like molecules to stably suppress gene expression^{4,5}. Here we report the construction of a set of retroviral vectors encoding 23,742 distinct shRNAs, which target 7,914 different human genes for suppression. We use this RNAi library in human cells to identify one known and five new modulators of p53-dependent proliferation arrest. Suppression of these genes confers resistance to both p53-dependent and p19^{ARF}-dependent proliferation arrest, and abolishes a DNA-damage-induced G1 cell-cycle arrest. Furthermore, we describe siRNA bar-code screens to rapidly identify individual siRNA vectors associated with a specific phenotype. These new tools will greatly facilitate large-scale loss-of-function genetic screens in mammalian cells.

RNAi is a defence mechanism triggered by double-stranded (ds)RNAs to protect cells from parasitic nucleic acids. The dsRNAs are processed into siRNAs, which target homologous RNAs for destruction⁶. In mammalian cells, an RNAi response can be triggered by 21-base-pair siRNAs, which can cause strong, but transient, inhibition of gene expression⁷. By contrast, vector-expressed shRNAs can suppress gene expression over prolonged periods^{4,5}. In the current study, we create a large set of vectors and use them to search for components of the p53 tumour-suppressor pathway. This pathway is crucial for genome integrity as it transmits both anti-proliferative and pro-apoptotic signals in response to a variety of stress signals⁸.

To construct a human RNAi library (the 'NKi library'), we selected 7,914 human genes for shRNA-mediated reduction in expression, known as knockdown. This collection of genes includes components of major cellular pathways, including the cell cycle, transcription regulation, stress signalling, signal transduction and important biological processes such as biosynthesis, proteolysis and metabolism. In addition, genes implicated in cancer and other diseases are included in the library (see Supplementary Table 1). To increase the likelihood of obtaining a significant inhibition of gene expression, we constructed three different shRNA vectors against each gene (23,742 vectors in total; Fig. 1a and Supplementary Fig. 1). The oligonucleotides specifying the shRNAs were annealed and cloned in a high-throughput fashion into pRetroSuper (pRS), a retroviral vector that contains the shRNA expression cassette⁹. Using a pool of three knockdown vectors against a single gene, we obtain on average 70% inhibition of expression for approximately 70% of the genes in the library (see ref. 10; data not shown). The vector-based shRNA library can be used for functional genetic screens in both short-term and long-term assays using DNA transfection or retroviral transduction.

To validate the RNAi library, we developed a cell system to screen for bypass of p53-dependent proliferation arrest. We generated primary human BJ fibroblasts, which ectopically express the murine ecotropic receptor, the telomerase catalytic subunit (TERT) and a temperature-sensitive allele of SV40 large T antigen (tsLT), yielding BJ-TERT-tsLT cells. As can be seen in Fig. 1b, these cells proliferate when grown at 32 °C, the temperature at which tsLT binds and inactivates both retinoblastoma protein (pRB) and p53, but enter into a synchronous proliferation arrest after a shift to 39 °C, at which tsLT is inactive. To determine whether this proliferation arrest is p53 dependent, we infected BJ-TERT-tsLT cells with pRS-p53 (which targets p53 for suppression) at 32 °C, and shifted them to 39 °C after two days. Figure 1b shows that knockdown of p53 allowed temperature-shift-induced proliferation arrest to be bypassed. Knockdown of the RB pathway component p16^{INK4A} alone did not allow growth arrest to be bypassed, but simultaneous suppression of both p16^{INK4A} and p53 yielded a further increase in escape from growth arrest compared with knockdown of p53 alone (Fig. 1b). Thus, the conditional proliferation arrest in BJ fibroblasts depends primarily on p53.

We isolated polyclonal plasmid DNA from each of the 83 pools of

NKi RNAi library vectors (Fig. 1a), and transfected this DNA into a packaging cell line to generate high-titre retroviral supernatants. These retroviruses were then used to infect BJ-TERT-tsLT cells at 32 °C. We co-infected cells with a shRNA vector that targets *p16^{INK4A}*, because suppression of this gene further enhances colony formation induced by *p53* knockdown (Fig. 1b). After two days, the cells were shifted to 39 °C and monitored for the appearance of proliferating colonies. In first-round screening, we identified six pools of infected cells that form colonies at 39 °C. To detect the biologically active shRNAs in these vector pools, we isolated several individual colonies of proliferating BJ fibroblasts and recovered the shRNA inserts of the vectors by polymerase chain reaction (PCR; see Methods). These shRNA inserts were then re-cloned, and their identity was established by DNA sequence analysis.

The majority of colonies of BJ fibroblasts that proliferated at 39 °C contained multiple shRNA inserts. Only those shRNA inserts that were present in multiple independently derived colonies were analysed in second-round selection in BJ-TERT-tsLT fibroblasts. Using this approach, we identified shRNAs against six genes that could suppress the temperature-shift-induced proliferation arrest in the BJ-TERT-tsLT cells in the second round. One of the six genes was *p53* itself, which underscores the quality of the NKi library. In

addition, we found that individual shRNAs against *RPS6KA6* (ribosomal S6 kinase 4, RSK4), *HTATIP* (histone acetyl transferase TIP60), *HDAC4* (histone deacetylase 4), *KIAA0828* (a putative S-adenosyl-L-homocysteine hydrolase, SAH3) and *CCT2* (T-complex protein 1, β -subunit) prevented the *p53*-dependent growth arrest in the BJ-TERT-tsLT fibroblasts (Fig. 1c). Knockdown of all five newly identified genes also mediated escape from proliferation arrest in BJ-TERT-tsLT cells when tested in the absence of *p16^{INK4A}* knockdown vector (Fig. 2b), suggesting that inhibition of these genes primarily allows the bypass of the *p53* response (see below).

Some siRNAs have 'off-target' effects, which are often the result of partial homology to other transcripts^{11,12}. The shRNA library was designed to avoid off-target effects by minimizing homology of shRNAs to other transcripts (see Methods). Furthermore, it is very unlikely that two independent siRNAs against the same transcript target a common off-target transcript for suppression. For three of the five newly identified genes, we found that two of the three shRNAs present in the library mediated escape from growth arrest in the BJ-TERT-tsLT cells, suggesting that the effects of the shRNAs were 'on-target'. For the two genes for which only one functional shRNA from the library was found, we designed a second active

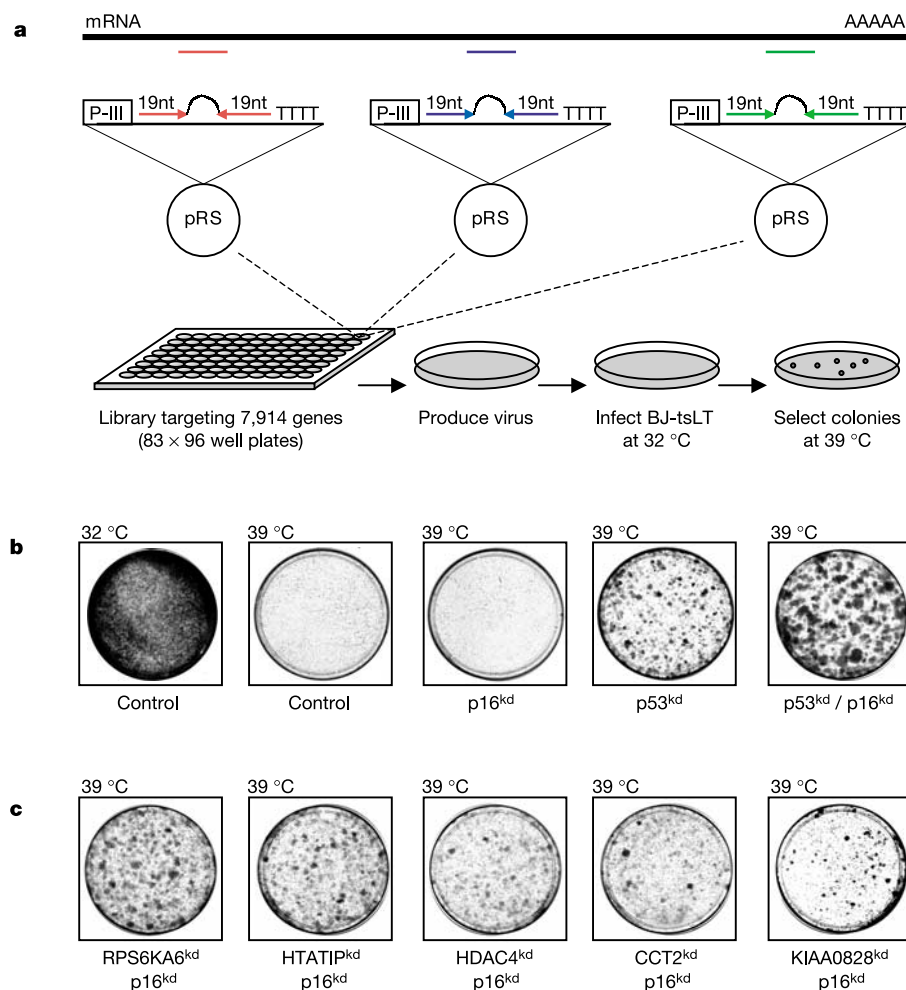


Figure 1 RNAi library screen in conditionally immortalized cells. **a**, Construction of the NKi RNAi library. From each gene transcript, three 19-nucleotide (nt) sequences were designed, and hairpin derivatives (59-mer oligonucleotides; see Supplementary Fig. 1) were cloned into pRS. Three vectors targeting one gene were combined in a single well of a 96-well plate. From each 96-well plate, DNA was pooled, and high-titre polyclonal virus was produced and used to infect BJ-tsLT cells at 32 °C. After two days, cells were shifted

to 39 °C, and after three weeks colonies were isolated and analysed. **b**, The growth arrest of temperature-shifted BJ-tsLT cells is *p53* dependent. Co-infection with an shRNA against *p16^{INK4A}* enhanced colony outgrowth. Cells were infected with the indicated shRNA knockdown (kd) vectors at 32 °C. Fifty thousand cells were seeded at both 32 °C and 39 °C, and colonies were stained after three weeks. **c**, Knockdown of the five newly identified genes allows bypass of the *p53*-dependent growth arrest in BJ-tsLT cells.

shRNA vector targeting the same transcript (see Supplementary Table 2). Figure 2a shows that transient co-transfection of each of the five pairs of two active shRNA vectors with the corresponding full-length complementary DNA expression vectors resulted in a substantial suppression of their cognate gene. Furthermore, we also observed a significant suppression of the corresponding endogen-

ous transcripts (see Supplementary Fig. 2). Moreover, all five pairs of two active shRNA vectors were able to rescue the temperature-shift-induced growth arrest in the BJ-TERT-tsLT cells (Fig. 2b). Together, these data suggest that the observed escape from growth arrest is due to on-target effects of the shRNA vectors on the corresponding genes.

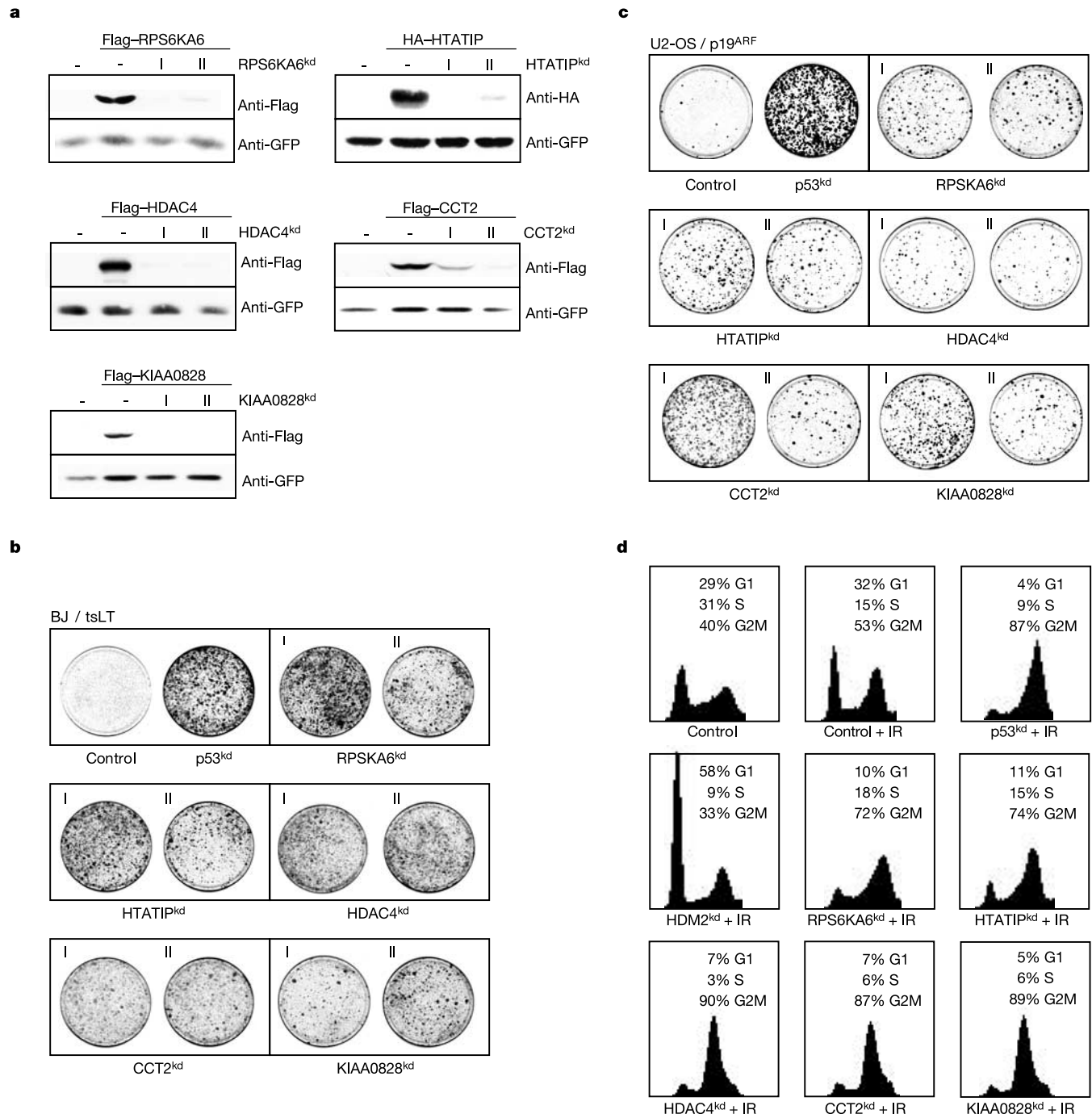


Figure 2 Validation of new p53 pathway components. **a**, Knockdown of newly identified p53 pathway components. U2-OS cells were co-transfected with the shRNA vectors (I or II) and their corresponding tagged (HA or Flag) cDNA expression vectors as indicated. Extracts were immunoblotted against Flag, HA or GFP (control). **b**, Knockdown of the five pairs of shRNA vectors rescue growth arrest in BJ-TERT-tsLT cells. Cells were infected with the indicated shRNA vectors (in the absence of the *p16^{INK4A}* knockdown vector), selected at 39°C, and colonies were stained after three weeks. **c**, Two independent

shRNA vectors of each indicated gene rescue a p19^{ARF}-induced growth arrest in U2-OS cells. Cells were co-infected with the shRNA vectors and LZRS-p19^{ARF}-IRES-zeocin. Fifty thousand cells were seeded and zeocin/puromycin selected. Colonies were stained after two weeks. **d**, Rescue of a p53-dependent cell-cycle arrest following ionizing radiation. U2-OS cells were co-transfected with the indicated shRNA vectors and H2B-GFP, irradiated (+IR, 10 Gy). GFP-positive cells were analysed by FACS.

The tumour suppressor p19^{ARF} elicits a p53-dependent proliferation arrest when ectopically expressed¹³. We argued that if the five newly identified genes act in the p53 pathway, their knockdown should confer resistance to p19^{ARF}-induced proliferation arrest. To test this, we used a retroviral vector encoding a p19^{ARF}-red-fluorescent-protein (ARF-RFP) fusion protein to infect human U2-OS osteosarcoma cells. Upon infection of U2-OS cells with this virus, only very few proliferating colonies appeared. The few colonies that appeared did not express the ARF-RFP fusion protein (Fig. 2c; data not shown). By contrast, many ARF-RFP-positive colonies appeared in U2-OS cells in which p53 was suppressed by shRNA (Fig. 2c). Importantly, each of the two independent shRNA vectors against the five newly identified genes allowed proliferation in the presence of ARF-RFP, which is shown by the presence of red fluorescence (Fig. 2c; data not shown). This result provides further support for the notion that the newly identified genes are components of the p53 pathway.

Next, we asked whether suppression of these genes allowed the p53-dependent G1 cell-cycle arrest to be bypassed after exposure to DNA damage. Figure 2d shows that exposure of U2-OS cells to ionizing radiation (IR) causes these cells to withdraw from S phase and arrest both in the G1 phase of the cell cycle (a p53-dependent process¹⁴) and in G2/M (a p53-independent process¹⁴). As expected, inactivation of p53 by shRNA completely abrogates the G1 fraction after IR, whereas introduction of a shRNA against HDM2 (the human orthologue of the mouse double minute 2 (*Mdm2*) gene), which activates p53, resulted in a marked increase in the G1 phase of the cell cycle (Fig. 2d). These results demonstrate that arrest in the G1 phase of the cell cycle after IR is determined by the activity of

p53. Transfection of knockdown vectors for the five newly identified genes in each case resulted in a cell-cycle distribution identical to the cells expressing the shRNA against p53 (Fig. 2d). Thus, the knockdown of all five genes prevents three different forms of p53-dependent proliferation arrest (Fig. 2b–d).

A major component of the anti-proliferative response of p53 is the CDK inhibitor p21^{cip1} (ref. 15), as it is a critical downstream component of the G1 arrest in response to DNA damage^{16,17} and is required for the senescence response of human fibroblasts¹⁸. To test whether the five newly identified genes affect the expression of known p53 target genes, we stably expressed knockdown vectors for all five genes in U2-OS cells and determined the expression of two p53 target genes, p21^{cip1} and BAX¹⁹. In addition, p53 protein levels can be used as an indirect read-out for p53 function, as p53 protein levels are controlled by the p53 target *HDM2* (refs 20, 21). Figure 3a shows that knocking down *RPS6KA6*, *HDAC4*, *CCT2*, *KIAA0828* or *HTATIP* causes a strong reduction in p21^{cip1} messenger RNA and protein levels, but does not affect p53 protein levels nor the levels of BAX mRNA and protein (Fig. 3a, b). Conversely, overexpression of *HTATIP* resulted in upregulation of p21^{cip1} expression (Fig. 3c) and further enhanced G1 cell-cycle arrest in response to DNA damage (Fig. 3d). We conclude that knockdown of all five new p53 pathway components leads to a specific downregulation of a subset of p53 target genes, which includes p21^{cip1}.

Next, we tested the significance of the inhibition of p21^{cip1} expression for the phenotypes induced by the newly identified genes. Figure 4a shows that suppression of p21^{cip1} was nearly as efficient at preventing growth arrest in BJ-TERT-tsLT cells as knockdown of p53 itself. Furthermore, knockdown of p21^{cip1}

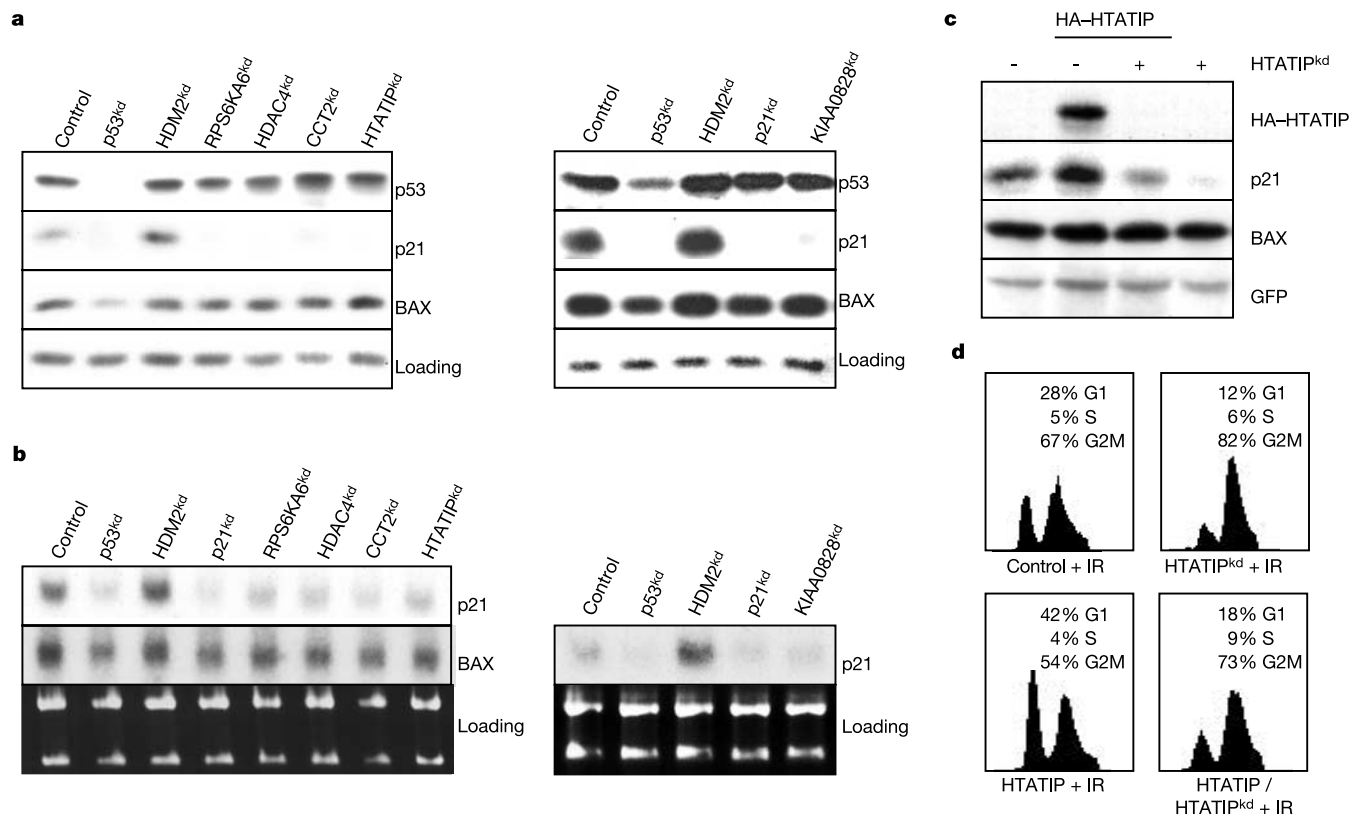


Figure 3 Downregulation of p21 transcription. **a**, U2-OS cells were stably infected with retroviruses expressing the indicated knockdown constructs. Cell extracts were immunoblotted with antibodies against p53, p21^{cip1} and BAX. **b**, mRNA levels of p21 and BAX in the stably infected U2-OS cells. As a loading control, RNA was stained with ethidium bromide. **c**, U2-OS cells were transfected with an expression vector for GFP and

CMV-HA-HTATIP/pRS-HTATIP as indicated. Cell extracts were immunoblotted with antibodies against HA, p21, BAX and GFP. **d**, Overexpression of HTATIP in U2-OS cells increased the population of cells in the G1 phase of the cell cycle after irradiation (+IR, 5 Gy). Co-transfection of a HTATIP shRNA vector reversed the effect of HTATIP overexpression.

allowed the p53-dependent growth arrest induced by p19^{ARF} (Fig. 4b) and the G1 cell-cycle arrest induced by IR in U2-OS cells (Fig. 4c) to be overridden. This suggests that suppression of p21^{cip1} is a major element in the cellular response to the effect of knocking down the newly identified p53 pathway components. However, we cannot exclude the possibility that knockdown of the newly identified genes has effects on other p53 target genes or p53-independent effects that explain (part of) their knockdown phenotype.

The RNAi screen described above is time consuming in that individual colonies of cells must be isolated and hairpin vectors recovered and tested in second-round selection. We have recently proposed an alternative strategy to rapidly screen complex shRNA vector libraries in a polyclonal format, a technique that we named 'siRNA bar-code screens'²². This approach, which was pioneered in yeast²³, takes advantage of the fact that each hairpin vector contains a unique gene-specific molecular identifier: the 19-mer targeting sequence. This 'molecular bar code' can be used to follow the relative abundance of individual vectors in a large population using DNA microarrays that contain the bar-code oligonucleotides²² (Fig. 5a). To test the concept of siRNA bar-code screens, we used a subset of the shRNA library (consisting of 990 different vectors) to infect U2-OS cells. After one day, genomic DNA was isolated from the infected cells, and the shRNA inserts were recovered by PCR with primers flanking the shRNA expression cassette (which includes the 19-mer gene-specific 'bar codes'). The PCR fragments were then labelled with fluorescent dye, and the relative abundance of specific shRNA inserts was determined by hybridization to a DNA microarray containing the 59-mer oligonucleotides used to generate the hairpin vectors (containing the bar-

code sequences). Figure 5a shows that the PCR fragments hybridize with high specificity to their complementary 59-mer oligonucleotides on the array.

To further validate the bar-code screening concept, we asked whether we could identify a mouse p53 shRNA vector mixed in a pool of 312 human shRNA vectors in a functional screen. We infected this set of shRNA vectors in conditionally immortalized mouse neuronal cells, which are immortal at 32 °C but enter into a synchronous p53-dependent growth arrest at 39 °C (ref. 24). Figure 5b shows the design of the experiment and the relative abundance of all the bar codes ten days after shift to 39 °C, which activated the p53-dependent growth arrest²⁴. Only the p53 knock-down vectors (indicated in red) were strongly selected in the population, whereas the set of control shRNA vectors was neither positively nor negatively selected. The positive selection of the p53 shRNA vectors occurred over time and only when the p53-dependent growth arrest was invoked (at 39 °C), and not at 32 °C, which is when p53 is inactive owing to LT binding (Fig. 5d). These data indicate that siRNA bar-code screening can be used to rapidly identify individual shRNA vectors that modulate cellular responses in large populations of vectors.

We identify here five components of the p53 tumour-suppressor pathway through a large-scale RNAi screen in mammalian cells. Inhibition of each of these five genes confers resistance to three different aspects of p53-dependent proliferation arrest, raising the possibility that these genes are tumour-suppressor genes. p53 is a central component of the cellular response to a variety of stress signals, and the genes identified here may be part of the machinery that allow cells to respond to such signals. Our data indicate that the

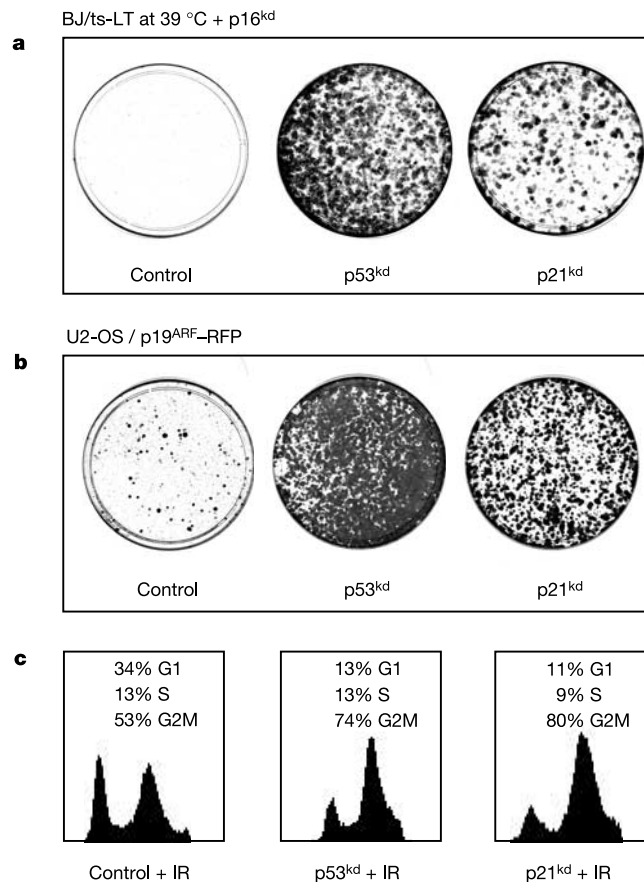


Figure 4 Knockdown of p21^{cip1} is sufficient to bypass a p53-dependent arrest. Introduction of p53 and p21 shRNA vectors rescue **a**, the growth arrest of

temperature-shifted BJ-tsLT cells, **b**, the p19^{ARF}-induced growth arrest in U2-OS cells and **c**, the p53-induced G1 arrest in U2-OS cells.

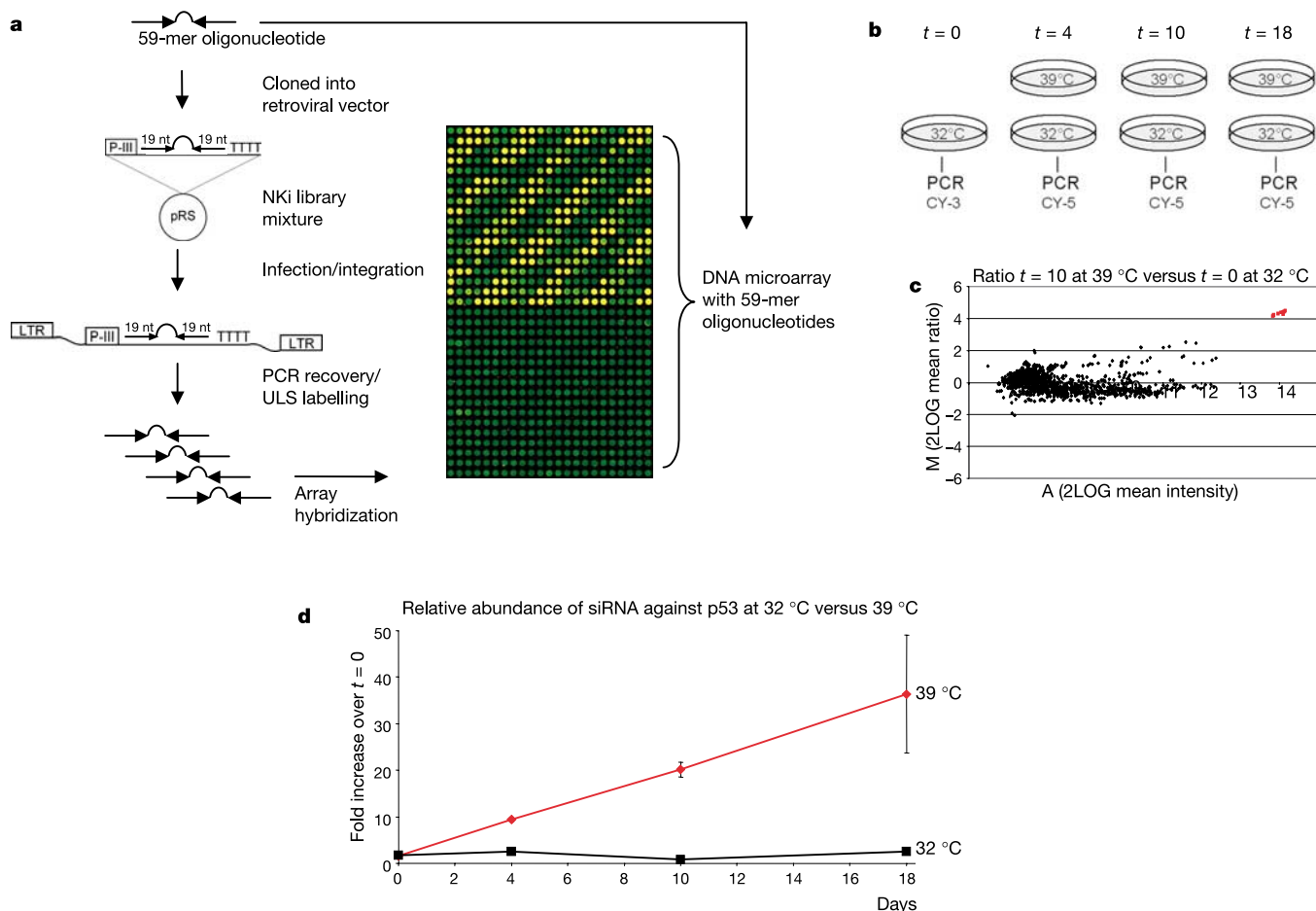


Figure 5 siRNA bar-code screens. **a**, DNA microarrays containing the 59-mer oligonucleotides present in a selected set of 990 different shRNA vectors were generated. Genomic DNA isolated from cells infected with these 990 shRNA vectors was used for PCR amplification. ULS-labelled bar-code fragments were hybridized on microarrays that consist of oligonucleotides that were present (upper half) or absent (lower half) in the selected set of 990 vectors. **b**, TsLT immortalized mouse neuronal cells²⁹ were infected with a retroviral supernatant containing 312 knockdown vectors directed against human genes and one against murine p53 (ref. 30). Cells were temperature shifted to induce a

senescence-like arrest and used for isolation of the bar codes at the indicated time points. **c**, Comparative hybridization of bar-code fragments isolated from the experiment described in **b** at $t = 0$ with bar-code fragments isolated ten days after the temperature shift. Red spots indicate the relative abundance of the p53 shRNA. Note that only the relative abundance of the p53 knockdown vector has increased. **d**, Analysis of the relative abundance of p53 shRNAs in tsLT-immortalized mouse neuronal cells recovered from cells grown at 32 °C, when endogenous p53 is inactivated by the tsLT antigen, or at 39 °C, when p53 is activated owing to loss of tsLT antigen expression.

newly identified genes have a role, either directly or indirectly, in modulating the activity of p53 on the *p21^{cip1}* promoter. We did not identify the shRNA for *p21^{cip1}* in the genetic screen, even though it was present in the RNAi library and active in *p21^{cip1}* suppression. This indicates that the first screen was not saturating and that additional p53 pathway components may be identified from this library.

Our data highlight the power of large-scale RNAi screens in mammalian cells. A potentially useful application of siRNA bar-code screens is in the identification of synthetic lethal interactions (a combination of two non-lethal mutations that together result in cell death). The identification of such genetic interactions in mammalian cells may facilitate the development of new and more specific classes of anticancer drugs. □

Methods

Materials, antibodies and vectors

IR was performed with a 2×415 Ci ¹³⁷Cs source. Antibodies against p21 (C-19), BAX (N-20), green fluorescent protein (GFP; FL), haemagglutinin (HA; Y-11) and CDK4 (C-22) were obtained from Santa Cruz, anti-Flag M2 from Sigma and anti-p53 (ab-7) from Oncogene Research Products. HA-tagged HTATIP was PCR amplified and cloned in pRC/CMV (Invitrogen). Flag-tagged RPS6KA6, CCT2 and KIAA0828 were generated by PCR

amplification and cloning into pcDNA3.1/HygroFlag. pRS, pRS-p16^{kd}, LZRS-p19^{ARF}-RFP-IRES-zeocin, Flag-HDAC4 and H2B-GFP have been described previously^{9,25,26}. pEGFP-N1 was obtained from Clontech. For the generation of knockdown vectors for p53, p21 and HDM2, the following 19-nucleotide sequences were used: p53, 5'-CTACATGTGTAACAGTTCC-3'; p21, 5'-GACCATGTGGACCTGTCAC-3'; and HDM2, 5'-GATGATGAGGTATATCAAG-3'. Control transfections were performed with a mixture of non-functional hairpins.

Oligonucleotide design and library construction

A representative mRNA sequence was selected for each target from UniGene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>). Sequences were masked to remove repetitive sequences using RepeatMasker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>), and vector contamination was masked by searching with NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) against UniVec (<http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html>). Three unique 19-mers for each target were selected, where possible, (1) to contain no stretches of four or more consecutive T or A residues (to avoid premature pol III transcription termination signals); (2) to have 30–70% overall GC content; (3) to lie within the coding sequence of the target gene; (4) to begin with a G or C residue (consistent with recently established rules for strand bias²⁷); (5) to begin after an AA dimer in the 5' flanking sequence; (6) to end just prior to a TT, TG or GT doublet in the 3' flanking sequence; (7) to not contain *Xho*I or *Eco*RI restriction-enzyme sites to facilitate subsequent shuttling of the knockdown cassette into other vector backbones; (8) to share minimal sequence identity with other genes; (9) to target all transcript variants represented by RefSeq mRNAs (<http://www.ncbi.nlm.nih.gov/RefSeq/>); and (10) to not overlap with other 19-mers selected from the same target sequence.

For cloning, pairs of complementary oligonucleotides were annealed and ligated into

HindIII/BglII-digested pRS in 96-well plates. Ligation reactions were transformed into competent DH5 α bacteria. Bacterial cultures were grown overnight and plasmid DNA was isolated. All manipulations were performed using a Multitask robot.

Additional information about the NKI RNAi library can be found at <http://screeninc.nki.nl/>.

Cell culture, transfection and retroviral infection

BJ fibroblasts were cultured in a 4:1 mixture of DMEM:M199 supplemented with 15% heat-inactivated fetal calf serum. U2-OS and Phoenix cells were cultured in DMEM supplemented with 10% heat-inactivated fetal calf serum. Transfections were performed with the calcium phosphate precipitation technique. Ecotropic retroviral supernatants were produced by transfection of Phoenix packaging cells. Viral supernatants were filtered through a 0.45 μ m filter, and infections were performed in the presence of 4 μ g ml⁻¹ polybrene (Sigma). Drug selections in U2-OS cells were performed with 2 μ g ml⁻¹ puromycin and 200 μ g ml⁻¹ zeocin (Invitrogen).

Recovery of shRNA inserts

Genomic DNA was isolated from expanded colonies using DNAzol (Life Technologies). PCR amplification of the shRNA inserts was performed with Expand Long Template PCR system (Roche) and the use of pRS-fw primer: 5'-CCCTTGAACCTCCTCGTTCGACC-3' and pRS-rev primer: 5'-GAGACGTGCTACTTCCATTGTG-3'. Products were digested with *EcoRI/XhoI* and recloned into pRS. Hairpins were sequenced with Big Dye Terminator (Perkin Elmer) using pRS-seq primer: 5'-GCTGACGTATCAACCCGCT-3'.

Cell-cycle analysis

For fluorescence-activated cell sorting (FACS) analysis, U2-OS cells were co-transfected with H2B-GFP to select for transfected cells. Forty-eight hours after transfection, cells were treated with IR (5 or 10 Gy). Twenty-four hours after IR, the cells were washed and fixed in 70% ethanol at 4 °C. Before FACS analysis, cells were washed with PHN (PBS, 20 mM HEPES, 0.5% NP40) and incubated with 10 μ g ml⁻¹ propidium iodide and 250 μ g ml⁻¹ RNase A. In each assay, 8,000 GFP-positive cells were collected by FACScan and analysed using the CellQuest program (Becton Dickinson).

Western and northern blotting

For western blots, puromycin-selected cells were lysed in RIPA buffer (50 mM Tris pH 8; 150 mM NaCl; 1% NP40; 0.5% DOC; 0.1% SDS). Thirty micrograms of protein was separated on 8–12% SDS-polyacrylamide gel electrophoresis and transferred to polyvinylidene difluoride membranes (Millipore). Western blots were probed with antibodies. For northern analysis, 15 μ g total RNA was loaded on a 1% agarose gel and blotted onto Hybond N+ (Amersham). ³²P-labelled probes for p21 and BAX were generated with PCR labelling.

siRNA bar-code screens

The shRNA inserts were amplified from genomic DNA by PCR using the primers pRS-fw primer 5'-CCCTTGAACCTCCTCGTTCGACC-3' and pRS8-rev primer 5'-TAAAGCGCATGCTCCAGACT-3'. PCR products were labelled with cyanine-3 or cyanine-5 fluorescent groups using the Universal Linkage System (ULS; Kreatech Biotechnology) and purified over a KreaPure (Kreatech Biotechnology) spin column as described previously²⁸. PCR products from two samples were combined and hybridized to oligonucleotide arrays in 40 μ l of 25% formamide, 5 $\times$ SCC, 0.01% SDS (containing poly d(A), yeast transfer RNA and COT-1 DNA). Samples were heated to 100 °C for 1 min and applied to the array. Samples were hybridized for 18 h at 42 °C, washed and scanned using an Agilent microarray scanner. Quantification of the resulting fluorescent images was performed with Image 5.6 (BioDiscovery); the local background was subtracted and the data normalized and ²log transformed. Additional information on bar-code screens can be found at <http://screeninc.nki.nl/>.

Received 11 November 2003; accepted 26 January 2004; doi:10.1038/nature02371.

- Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Ashrafi, K. *et al.* Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272 (2003).
- Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
- Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
- Sherr, C. J. The Ink4a/ARF network in tumour suppression. *Nature Rev. Mol. Cell Biol.* **2**, 731–737 (2001).
- Brummelkamp, T., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243–247 (2002).
- Brummelkamp, T. R., Nijman, S. M., Dirac, A. M. & Bernards, R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF- κ B. *Nature* **424**, 797–801 (2003).
- Semizarov, D. *et al.* Specificity of short interfering RNA determined through gene expression signatures. *Proc. Natl Acad. Sci. USA* **100**, 6347–6352 (2003).
- Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
- Kamijo, T. *et al.* Tumor suppression at the mouse Ink4a locus mediated by the alternative reading frame product p19ARF. *Cell* **91**, 649–659 (1997).

- Lakin, N. D. & Jackson, S. P. Regulation of p53 in response to DNA damage. *Oncogene* **18**, 7644–7655 (1999).
- El-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Brugarolas, J. *et al.* Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature* **377**, 552–557 (1995).
- Waldman, T., Kinzler, K. W. & Vogelstein, B. p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res.* **55**, 5187–5190 (1995).
- Brown, J. P., Wei, W. & Sedivy, J. M. Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. *Science* **277**, 831–834 (1997).
- Miyashita, T. & Reed, J. C. Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. *Cell* **80**, 293–299 (1995).
- Barak, Y., Juven, T., Haffner, R. & Oren, M. mdm2 expression is induced by wild type p53 activity. *EMBO J.* **12**, 461–468 (1993).
- Kubbutat, M. H., Jones, S. N. & Vousden, K. H. Regulation of p53 stability by Mdm2. *Nature* **387**, 299–303 (1997).
- Brummelkamp, T. R. & Bernards, R. New tools for functional mammalian cancer genetics. *Nature Rev. Cancer* **3**, 781–789 (2003).
- Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450–456 (1996).
- Brummelkamp, T. R. *et al.* TBX-3, the gene mutated in Ulnar-Mammary Syndrome, is a negative regulator of p19ARF and inhibits senescence. *J. Biol. Chem.* **277**, 6567–6572 (2002).
- Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8**, 377–385 (1998).
- Wang, A. H. *et al.* HDAC4, a human histone deacetylase related to yeast HDAC1, is a transcriptional corepressor. *Mol. Cell Biol.* **19**, 7816–7827 (1999).
- Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
- Heetebrij, R. J. *et al.* Platinum(II)-based coordination compounds as nucleic acid labeling reagents: synthesis, reactivity, and applications in hybridization assays. *ChemBiochem* **4**, 573–583 (2003).
- Trettel, F. *et al.* Dominant phenotypes produced by the HD mutation in STHdh(Q111) striatal cells. *Hum. Mol. Genet.* **9**, 2799–2809 (2000).
- Dirac, A. M. & Bernards, R. Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. *J. Biol. Chem.* **278**, 11731–11734 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Friend and J. Downward for their support of this project, M. Voorhoeve, Z. Wu, X.-j. Yang, H. Yntema and Kreatech Biotechnology for reagents, the NKI microarray facility group for assistance, A. Dirac and S. Nijman for technical help, and members of the Bernards laboratory for discussions. This work was supported by grants from the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO), Cancer Research UK (CRUK), the Centre for Biomedical Genetics (CBG), the Dutch Cancer Society (KWF) and Utrecht University (ABC cluster).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.L.B. (r.beijersbergen@nki.nl) or R.B. (r.bernards@nki.nl).

Functional interactions between receptors in bacterial chemotaxis

Victor Sourjik* & Howard C. Berg

¹*Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, and the Rowland Institute at Harvard, 100 Edwin H. Land Boulevard, Cambridge, Massachusetts 02142, USA*

* Present address: ZMBH, University of Heidelberg, Im Neuenheimer Feld 282, D-69120, Germany

Bacterial chemotaxis is a model system for signal transduction, noted for its relative simplicity, high sensitivity, wide dynamic range and robustness. Changes in ligand concentrations are sensed by a protein assembly consisting of transmembrane receptors, a coupling protein (CheW) and a histidine kinase (CheA)^{1–4}. In *Escherichia coli*, these components are organized at the cell poles in tight clusters that contain several thousand copies of each protein^{1,4–6}. Here we studied the effects of variation

in the composition of clusters on the activity of the kinase and its sensitivity to attractant stimuli, monitoring responses *in vivo* using fluorescence resonance energy transfer. Our results indicate that assemblies of bacterial chemoreceptors work in a highly cooperative manner, mimicking the behaviour of allosteric proteins. Conditions that favour steep responses to attractants in mutants with homogeneous receptor populations also enhance the sensitivity of the response in wild-type cells. This is consistent with a number of models^{7–11} that assume long-range cooperative interactions between receptors as a general mechanism for signal integration and amplification.

The regulation of the autokinase activity of CheA has been described in terms of a two-state model in which a receptor is in either an active or an inactive state, either promoting or inhibiting the activity of CheA^{12–15}. Binding of attractant increases the probability that a receptor is inactive, whereas methylation of a receptor on four specific glutamate residues increases the probability that it is active^{3,4,16}. Receptors in lower modification states, although less active, have higher affinities to attractant^{15–18}. In its simple form, the two-state model is unable to explain many features of the chemotactic response, such as high sensitivity and integration of different stimuli. *Escherichia coli* has five types of receptor: two high-abundance receptors, for aspartate (Tar) and serine (Tsr), and three low-abundance receptors, Tap, Trg and Aer. There is growing evidence of interactions between receptors in the cell^{19,20}, and modifications of the two-state model have been proposed to account for signal amplification and integration based on receptor–receptor interactions that induce long-range cooperative behaviour^{7–11}. *In vitro* studies of receptor cooperativity have shown both high¹⁸ and low^{21,17}

cooperativity. Here we explore interactions between receptors *in vivo*.

We used an assay based on fluorescence resonance energy transfer (FRET) to monitor the activity of the kinase CheA in a cell population. This assay relies on phosphorylation-dependent interactions of the response regulator CheY, fused to yellow fluorescent protein (YFP), with its phosphatase CheZ, fused to cyan fluorescent protein (CFP). Under the conditions of our experiment, CheY is phosphorylated by CheA at the same rate at which CheY-P is dephosphorylated by CheZ, so the kinase activity can be inferred from the concentration of the CheZ–CheY-P enzyme–substrate complex and expressed in units of FRET; that is, as a function of the degree to which excitation of CFP generates fluorescence of YFP²². We added and subtracted attractant over a range of concentrations and constructed dose–response curves that, for convenience, were fit by the Hill equation, yielding a sensitivity, $K_{1/2}$, characterized by the ligand concentration at half-maximal response, $K_{1/2}$, and a cooperativity characterized by the Hill coefficient, H (see Supplementary Information). These experiments were repeated with cells in which the levels of expression of Tar, Tsr, Tap, CheA and/or CheW were varied (with the levels of Trg and Aer unchanged). Unless specified otherwise, we used strains lacking the enzymes required for adaptation: the methyltransferase CheR and the methyl-erastase CheB. In these strains receptors encoded by wild-type genes remain in the half-modified state, QE_{QE}, where Q (glutamine) mimics methylated glutamate, and receptors engineered in other modification states remain as such. Thus, all receptor molecules in *cheR cheB* cells are either identical or are of a few defined types.

We observed a strong effect of receptor homogeneity on both the sensitivity and cooperativity of the response (Fig. 1). In cells expressing native levels of Tar and Tsr (Fig. 1a), only a fraction of the CheA activity was Tar-dependent. This activity was inhibited by addition of the non-metabolizable aspartate analogue α -methyl-

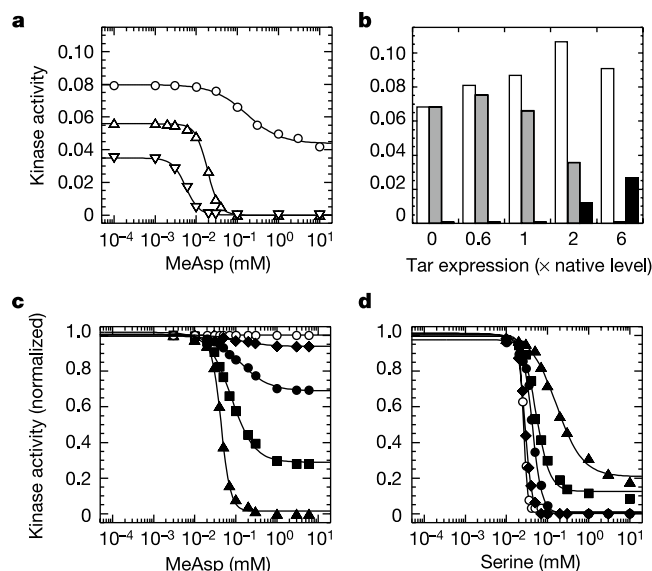


Figure 1 Effect of the homogeneity of the receptor cluster on response to attractants. **a**, Response to MeAsp of Tar⁺ Tsr⁺ Tap⁺ cells (circles), Tar⁺ Tap⁺ cells (triangles), or Tar⁺ cells (inverted triangles). **b**, Effect of level of expression of Tar in Tsr⁺ cells on the kinase activity in the absence of attractant or in the presence of a saturating dose of MeAsp (5 mM) or serine (10 mM). In all cases, a saturating dose of both attractants reduced the kinase activity to zero (not shown). White bars, no attractant; grey bars, MeAsp; black bars, serine. **c**, **d**, Response of cells expressing different levels of Tar in Tsr⁺ cells to MeAsp (**c**) or serine (**d**), normalized to the total pre-stimulus kinase activity (shown as white bars in **b**). The expression levels of Tar were 0 (open circles), 0.6 (diamonds), 1 (filled circles), 2 (squares) or 6 (triangles) times the native level. Here and in other figures kinase activity is expressed in units of FRET and solid lines are fits using the Hill equation. The fits and the best-fit values are summarized in Supplementary Information.

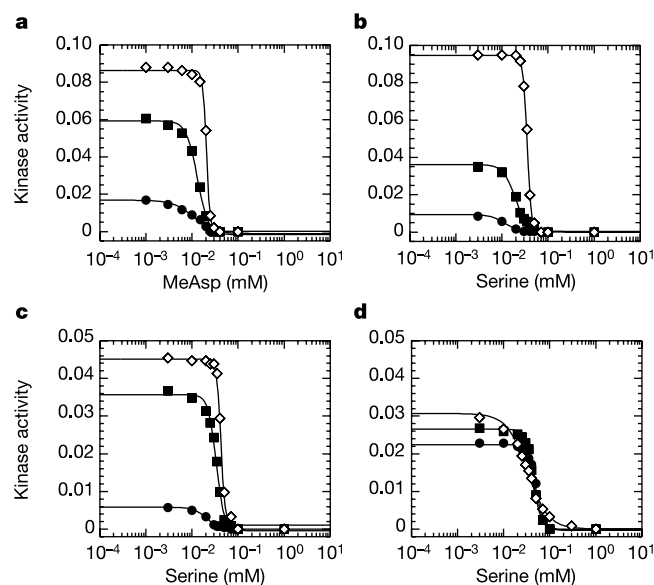


Figure 2 Effect of the composition of the receptor–kinase complex on response to attractants. **a**, Response to MeAsp of cells expressing only Tar at 1 (filled circles), 2 (filled squares) and 6 (open diamonds) times the native level. **b**, Response to serine of cells expressing only Tsr at 0.3 (filled circles), 0.7 (filled squares) and 5 (open diamonds) times the native level. **c**, Response to serine of Tsr⁺ cells expressing CheW at 0.01 (filled circles), 0.1 (filled squares) and 0.7 (open diamonds) times the native level. **d**, Response to serine of Tar⁺ Tsr⁺ Tap⁺ cells expressing CheA at 0.25 (filled circles), 0.3 (filled squares) and 8 (open diamonds) times the native level.

aspartate (MeAsp), with a half-maximal response at $K_{1/2}$ of 160 μM . In contrast, the entire kinase activity could be inhibited by serine. Deletion of *tsr* not only made the entire CheA activity Tar-dependent, but also resulted in an approximately eightfold increase in sensitivity towards MeAsp. This increase in sensitivity was accompanied by a significant increase in the cooperativity of the response, shifting the Hill coefficient from ~ 1 to ~ 3 . Additional deletion of a minor receptor, *tap*, resulted in a further roughly threefold increase in sensitivity but little change in cooperativity.

We obtained further evidence that Tar and Tsr affect each other's activity and sensitivity to attractants by varying the level of expression of Tar in a strain that was wild type for *tsr* (Fig. 1b–d). Figure 1b shows the kinase activity as a function of the level of expression of Tar before and after addition of a saturating dose of MeAsp or serine. The responses to these ligands were not additive. At low expression levels of Tar, the dominant receptor was Tsr, and the addition of serine reduced the kinase activity to zero. With Tar at six times the native level, Tar became the dominant receptor. When expression levels of Tar were increased from 0 (or 0.6) to six times the wild-type level, the sensitivity to MeAsp increased about threefold (Fig. 1c) and that to serine decreased about sixfold

(Fig. 1d), whereas the Hill coefficients for the response to MeAsp increased from ~ 1 to ~ 4 (Fig. 1c) and the Hill coefficients for the response to serine decreased from ~ 9 to ~ 1 (Fig. 1d). These data suggest that receptors interact with one another in clusters—the larger the relative number of receptors of a given kind in the complex, the higher the sensitivity and cooperativity of the response to the ligand for that receptor.

We found that clusters of receptors of essentially one kind gave responses of high cooperativity, depending on the receptor to CheW to CheA ratio (Fig. 2). Increasing the level of expression of either Tar or Tsr in a *tsr tar tap* strain gave responses to MeAsp or serine with Hill coefficients increasing from ~ 2 to ~ 10 , respectively (Fig. 2a, b). Similarly, increasing the level of expression of CheW in a *tar tap* strain raised the cooperativity of the response to serine, from a Hill coefficient of ~ 3 to ~ 10 (Fig. 2c). In contrast, increasing the level of expression of CheA lowered the cooperativity of the response to serine, from a Hill coefficient of ~ 5 to ~ 2 (Fig. 2d). Co-expression of both CheA and CheW at different induction levels did not affect the cooperativity of the response (data not shown).

High cooperativity was not a result of abnormally high levels of protein expression, because responses to serine with Hill coefficients

Box 1

MWC-type models of receptor cooperativity

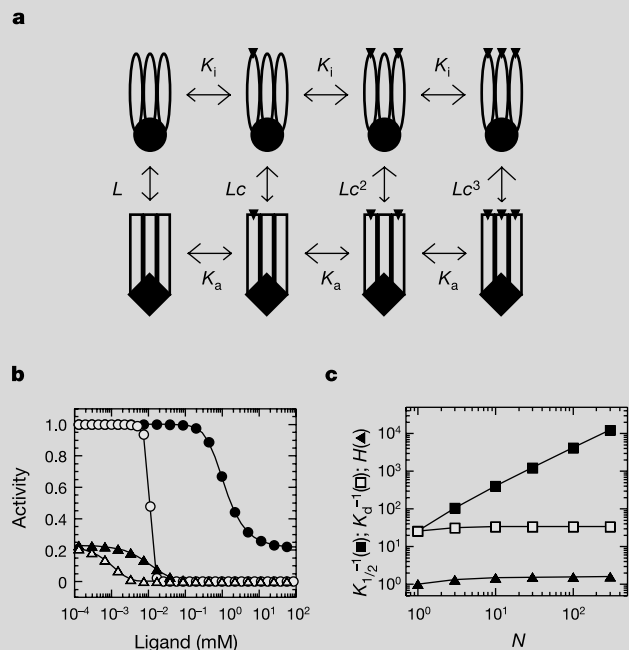
The MWC model²⁷ can be seen as an extension of the two-state model of chemoreceptors^{12–15} that includes interactions between receptors.

There are two crucial assumptions: (1) the inactive state of a receptor homodimer has a higher affinity to attractant than the active state; and (2) the entire complex exists with all of its N receptor homodimers either in active or inactive states. As in the two-state model, the receptor and the kinase are tightly coupled, so that the kinase is active when the receptor is active. Cooperativity of the response in this model depends not only on N but on two other parameters: L , the ratio of the probability that the complex is in the active state to the probability that it is in the inactive state in the absence of attractant (that is, the equilibrium constant), and c , the ratio of dissociation constants of attractant for inactive and active states of receptors, $K_i/K_a < 1$. Panel **a** shows a complex consisting of three receptor homodimers ($N = 3$), either inactive (ovals) or active (rectangles), associated with one CheA dimer, either inactive (circle) or active (diamond). CheW, another component of the complex, is not shown. On binding of attractant (small filled triangle) to one of the receptors, the equilibrium for the entire complex shifts towards the inactive state. Because receptors in the inactive state have higher affinity to attractant, the probability that the other receptors become occupied increases, further stabilizing the inactive state. Such positive reinforcement between attractant binding and receptor inactivation results in a cooperative response to attractant. The model can be extended to any number N of interacting receptor homodimers. Cooperativity (the Hill coefficient, H) increases with N and L , whereas sensitivity (the inverse of the concentration at half response, $K_{1/2}^{-1}$) increases with N but decreases with L . Panel **b** shows a simulation of the response of a receptor cluster to ligand at two values of N —3 (filled symbols) and 30 (open symbols)—and two values of L —0.3 (triangles) and 10^4 (circles). Here $K_i = 0.03 \text{ mM}$ and $K_a = 1 \text{ mM}$ (see Supplementary Information). If L is kept constant and low, the cooperativity remains approximately constant, whereas the sensitivity increases as $\sim N$. Ligand binding, characterized by K_d (the concentration of ligand at which half the receptors are occupied), also remains approximately constant. Panel **c** shows how values of $K_{1/2}^{-1}$, K_d^{-1} and H change with increasing number of interacting receptors at a constant $L = 0.3$, with values for K_i and K_a as in **b**.

The classical MWC model assumes that the receptor complex consists of N identical subunits with the same microscopic ligand-

binding affinities. In a mixed complex with the same total number of receptors but of different types, the cooperativity will be lower and will depend on the ratio between the number of different receptor species and their respective affinities to the ligand.

The assumption that the entire complex must 'flip' between the two states in a concerted fashion can be relaxed by assuming a finite coupling strength between receptors, as in the model of an extended two-dimensional receptor lattice proposed by refs 9, 10. A receptor cluster of $R \approx 10^3$ receptors with a receptor complex size $N \approx 10$ can then be thought of as consisting of either R/N independent all-or-none complexes or of one extended lattice in which the typical distance of signal propagation is $\sim N^{1/2}$ and the effective number of interacting receptors is N .



of ~ 10 were observed at native levels of Tsr, CheA and CheW (Figs 1d and 2c). Nor was high cooperativity an artefact due to saturation of the FRET response, because it was observed at pre-stimulus levels of 0.05–0.07, levels much lower than the highest FRET signals of 0.11 measured in our experiments (Figs 1b, d and 2c). Rather, the high cooperativity must reflect extensive interactions between receptors in native clusters. Several models have been proposed to explain the structure of receptor clusters^{20,23–25}. Interest has focused on the structure consisting of two trimers of homodimers (that is, six receptor dimers) associated with one CheA dimer^{20,24,26}; however, our data suggest cooperative interactions that involve more than ten receptor homodimers and thus require higher levels of organization, with a variable stoichiometry of the receptor complex.

It is possible to understand these results in the context of the classical Monod–Wyman–Changeux (MWC) model describing behaviour of multi-subunit allosteric proteins²⁷ (Box 1). With this model, fits to the data in both Figs 1 and 2 are virtually identical to those obtained with the Hill equation (comparisons not shown). The major effect of increasing the homogeneity of Tar on deletion of *tsr* and *tap* (Fig. 1a) is to increase the number (N) of interacting Tar homodimers in the complex (see Supplementary Information for parameters of the fits). In contrast, the major effect of increasing levels of expression of Tar in *tar tsr tap* cells (Fig. 2a) is to increase L ; that is, to increase the probability that the receptor–kinase complex is in the active state in the absence of attractant. The best-fit values of N increased as well. Similar results were obtained for changes in levels of expression of Tsr (data not shown). Although exact best-fit values of N and L were dependent on the values assumed for dissociation constants of attractants for inactive and active receptor states (see Supplementary Information), N always was greater than the Hill coefficient of the fit to the same data.

Receptor interactions provide a mechanism for integrating the signal between receptors of different types, as well as between receptors in different modification states. When Tar receptors in singly modified (QEEE) and triply modified (QEQQ) states were

co-expressed in the same cells, neither the initial activity nor the sensitivity to MeAsp was a simple combination of the results obtained when these receptors were expressed separately (Fig. 3a). In addition, a strain co-expressing Tsr(QEQE) with Tar(EEEE) showed lower kinase activity and was approximately 3.3 times more sensitive to serine than a strain co-expressing Tsr(QEQE) and Tar(QEQQ) (Fig. 3b). Conversely, when the activity of Tsr receptors was reduced by $\sim 50\%$ by the addition of serine in a strain expressing both Tsr and Tar in a QEQE state, sensitivity of Tar receptors to MeAsp increased roughly 4.5-fold (Fig. 3c). All three effects presented in Fig. 3 can be explained by analogy to the MWC model (Box 1). Receptors that are less active because they are in a low modification state and/or bound to attractant will decrease the activity and increase the sensitivity of receptors with which they interact, regardless of their type. Receptors that are more active because they are in a high modification state and/or attractant-free will have the opposite effect.

Conditions that favoured high cooperativity in *cheR cheB* cells favoured high sensitivity in wild-type ($\text{CheR}^+ \text{CheB}^+$) cells (Fig. 4). When expression levels of Tar were increased in wild-type cells, the sensitivity to MeAsp increased about fivefold, whereas the Hill coefficient (~ 2) remained unchanged (Fig. 4a). Similar results were obtained when expression levels of CheA were decreased (Fig. 4b) or when expression levels of CheW were increased (data not shown). $\text{CheR}^+ \text{CheB}^+$ cells have receptors with a range of modification states mixed in each receptor complex. The MWC model can be applied to mixed complexes if the ratio between inactive and active receptor states changes with receptor modification but dissociation constants of attractant for these two states remain constant¹⁵. Increasing levels of Tar expression in $\text{CheR}^+ \text{CheB}^+$ cells increases the best-fit values of N , whereas the values of L remain roughly constant (Fig. 4c). In *cheR cheB* cells, with homogeneous receptor populations, the best-fit values of N increase in a similar fashion, but the values of L increase by several orders of magnitude (Fig. 4c). Evidently, the methylation system in $\text{CheR}^+ \text{CheB}^+$ cells offsets this increase by adjusting the fraction of

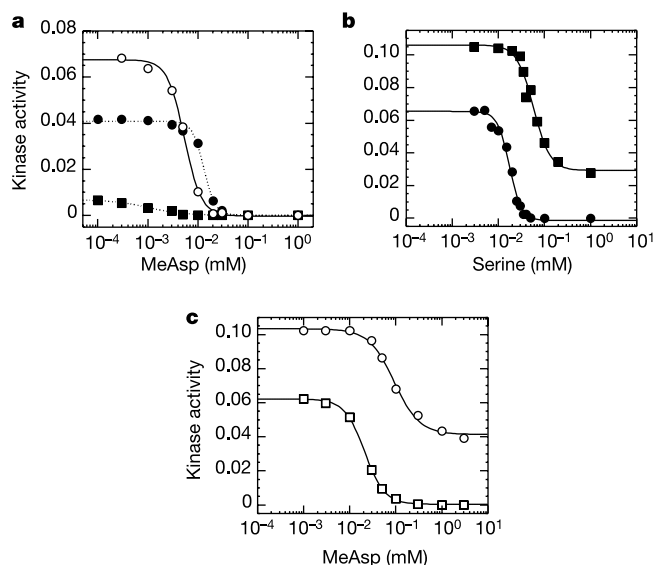


Figure 3 Cross-talk between receptors in different modification states and between different types of receptor. **a**, Response to MeAsp of cells expressing Tar(QEEE) at ~ 1.5 times the native level (filled squares), Tar(QEQQ) at the native level (filled circles), or both at these levels (open circles). **b**, Response to serine of Tsr^+ cells expressing Tar(EEEE) (circles) or Tar(QEQQ) (squares) at ~ 1.5 times the native level. **c**, Response to MeAsp of Tsr^+ cells expressing Tar(QEQE) at ~ 2 times the native level, in the presence (squares) or absence (circles) of 0.07 mM serine.

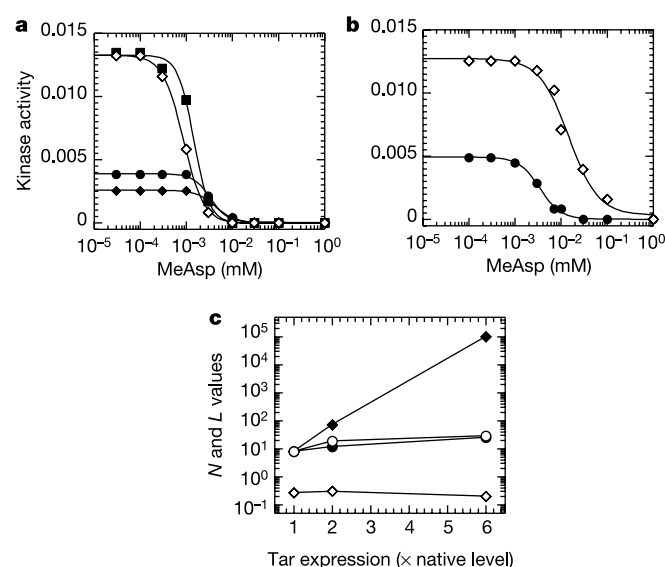


Figure 4 Effect of complex stoichiometry on response of $\text{CheR}^+ \text{CheB}^+$ cells to attractant. **a**, Response to MeAsp of cells expressing only Tar at 0.6 (filled diamonds), 1 (filled circles), 2 (filled squares) and 6 (open diamonds) times the native level. **b**, Response to MeAsp of $\text{Tar}^+ \text{Tsr}^+ \text{Tap}^+$ cells expressing CheA at 0.25 (circles) and 8 (diamonds) times the native level. **c**, Best-fit values of N (circles) and L (diamonds) for fits of the MWC model to the data in Fig. 2a (filled symbols) and Fig. 4a (open symbols).

receptors in high modification states to keep the activity of the receptor complex at a moderate level (~ 0.3). In the MWC model, large Hill coefficients are observed only when values of both N and L are large, whereas high sensitivity is observed when N is large and L is small (Box 1). Thus, high sensitivity of wild-type cells can be explained by combination of receptor interactions and control over receptor activity. CheB is particularly crucial²² because it generates lower modification states with lower activity and higher sensitivity. Clearly, more remains to be learned about the details of signal processing and amplification in chemotaxis, and more complicated models are required to quantitatively account for integration of signal between receptors of different types; however, the functional interactions between receptors observed here provide a framework for further experimental analysis and modelling. □

Methods

Plasmids

The genes *cheY-yfp* and *cheZ-cfp*, encoding CheY-YFP and CheZ-CFP fusions²², were amplified with polymerase chain reaction (PCR) and cloned in tandem, using PCR-introduced *SacI*, *XbaI* and *HindIII* restriction sites, into pTrc99A (Amp^R, Pharmacia) under an isopropyl β -D-thiogalactoside (IPTG)-inducible promoter, yielding pVS88. All other chemotaxis proteins were expressed under the tight control of a salicylate-inducible promoter from the pLC112-derived plasmids²⁰. The Tar expression plasmid pLC113, the Tsr expression plasmid pPA114, the CheA expression plasmid pPA113 and the pLC112-derived expression vector pKG110 were obtained from J. S. Parkinson. Plasmids expressing CheA and CheW or only CheW (pVS125 and pVS126, respectively) were constructed by cloning PCR-amplified *cheA* and *cheW* or *cheW* genes into *NdeI* and *KpnI* restriction sites of pKG110. Plasmid pVS121 that expresses Tar(Q295E/Q309E) (referred to as Tar(EEEE)), pVS120 that expresses Tar(Q309E) (Tar(QEEE)) and pVS122 that expresses Tar(E491Q) (Tar(QEQQ)) were constructed by replacing the *KpnI/BamHI* fragment of pLC113 with a fragment containing corresponding mutations of *tar* that were amplified via PCR from strains VS134, VS131 and VS147 (ref. 22), respectively.

Strains and their growth

Derivatives of *E. coli* K12 strain RP437 used in this study are listed in Supplementary Information. Ordinarily, proteins were expressed from plasmids in strains deleted for the chromosomal copy of the corresponding gene. Cells were grown as described²² in the presence of 100 $\mu\text{g ml}^{-1}$ ampicillin, 34 $\mu\text{g ml}^{-1}$ chloramphenicol, 50 μM IPTG and varying amounts (0–1 μM) of sodium salicylate. Immunoblots, performed as before⁶, were used to compare the expression levels of proteins induced by salicylate with their native expression levels. Polyclonal antibody raised against the signalling domain of Tsr¹³ that reacts equally well with both Tar and Tsr was obtained from J. S. Parkinson. Polyclonal antibodies raised against CheA and CheW were obtained from R. C. Stewart.

FRET measurements

Cells expressing CheY-YFP and CheZ-CFP were attached to a polylysine-coated coverslip and placed in a flow cell. The flow cell was kept under a constant flow (0.5 ml min⁻¹) of tethering buffer (10 mM potassium phosphate, 0.1 mM EDTA, 1 μM L-methionine, 10 mM sodium lactate, pH 7). Solutions of the same buffer were used to add and remove specified amounts of attractants (α -methyl-D,L-aspartate or L-serine) in a sequence of steps of increasing size. FRET, defined as the fractional change in cyan fluorescence due to energy transfer, was calculated from changes in the ratios of yellow and cyan fluorescence signals, measured as described before^{22,28}. For CheR⁺ CheB⁺ cells, the initial response was measured (before the cells had time to adapt). A field of 300–500 cells was monitored in each experiment. Under the conditions of our experiments, interaction between CheY-YFP and CheZ-CFP provides a measure of the activity of CheA²², so FRET data were plotted as kinase activity. Zero CheA activity was determined by adding saturating amounts of serine and MeAsp.

Data analysis and simulations

Data analysis and simulations are described in Supplementary Information.

Received 19 November 2003; accepted 5 February 2004; doi:10.1038/nature02406.

- Gegner, J. A., Graham, D. R., Roth, A. F. & Dahlquist, F. W. Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**, 975–982 (1992).
- Boukhvalova, M. S., Dahlquist, F. W. & Stewart, R. C. CheW binding interactions with CheA and Tar. Importance for chemotaxis signaling in *Escherichia coli*. *J. Biol. Chem.* **277**, 22251–22259 (2002).
- Borkovich, K. A., Kaplan, N., Hess, J. F. & Simon, M. I. Transmembrane signal transduction in bacterial chemotaxis involves ligand-dependent activation of phosphate group transfer. *Proc. Natl Acad. Sci. USA* **86**, 1208–1212 (1989).
- Ninfa, E. G., Stock, A., Mowbray, S. & Stock, J. B. Reconstitution of the bacterial chemotaxis signal transduction system from purified components. *J. Biol. Chem.* **266**, 9764–9770 (1991).
- Maddock, J. R. & Shapiro, L. Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**, 1717–1723 (1993).
- Sourjik, V. & Berg, H. C. Localization of components of the chemotaxis machinery of *Escherichia coli* using fluorescent protein fusions. *Mol. Microbiol.* **37**, 740–751 (2000).

- Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
- Shi, Y. & Duke, T. Cooperative model of bacterial sensing. *Phys. Rev. E* **58**, 6399–6406 (1998).
- Duke, T. A. J. & Bray, D. Heightened sensitivity of a lattice of membrane receptors. *Proc. Natl Acad. Sci. USA* **96**, 10104–10108 (1999).
- Shimizu, T. S., Aksenov, S. V. & Bray, D. A spatially extended stochastic model of the bacterial chemotaxis signalling pathway. *J. Mol. Biol.* **329**, 291–309 (2003).
- Mello, B. A. & Tu, Y. Quantitative modeling of sensitivity in bacterial chemotaxis: the role of coupling among different chemoreceptor species. *Proc. Natl Acad. Sci. USA* **100**, 8223–8228 (2003).
- Asakura, S. & Honda, H. Two-state model for bacterial chemoreceptor proteins. The role of multiple methylation. *J. Mol. Biol.* **176**, 349–367 (1984).
- Ames, P. & Parkinson, J. S. Constitutively signaling fragments of Tsr, the *Escherichia coli* serine chemoreceptor. *J. Bacteriol.* **176**, 6340–6348 (1994).
- Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
- Morton-Firth, C. J., Shimizu, T. S. & Bray, D. A. Free-energy-based stochastic simulation of the Tar receptor complex. *J. Mol. Biol.* **286**, 1059–1074 (1999).
- Borkovich, K. A., Alex, L. A. & Simon, M. I. Attenuation of sensory receptor signaling by covalent modification. *Proc. Natl Acad. Sci. USA* **89**, 6756–6760 (1992).
- Bornhorst, J. A. & Falke, J. J. Attractant regulation of the aspartate receptor-kinase complex: Limited cooperative interactions between receptors and effects of the receptor modification state. *Biochemistry* **39**, 9486–9493 (2000).
- Li, G. & Weis, R. M. Covalent modification regulates ligand binding to receptor complexes in the chemosensory system of *Escherichia coli*. *Cell* **100**, 357–365 (2000).
- Gestwicki, J. E. & Kiessling, L. L. Inter-receptor communication through arrays of bacterial chemoreceptors. *Nature* **415**, 81–84 (2002).
- Ames, P., Studdert, C. A., Reiser, R. H. & Parkinson, J. S. Collaborative signaling by mixed chemoreceptor teams in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 7060–7065 (2002).
- Levit, M. N. & Stock, J. B. Receptor methylation controls the magnitude of stimulus-response coupling in bacterial chemotaxis. *J. Biol. Chem.* **277**, 36760–36765 (2002).
- Sourjik, V. & Berg, H. C. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **99**, 123–127 (2002).
- Shimizu, T. S. *et al.* Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. *Nature Cell Biol.* **2**, 792–796 (2000).
- Levit, M. N., Grebe, T. W. & Stock, J. B. Organization of the receptor-kinase signaling array that regulates *Escherichia coli* chemotaxis. *J. Biol. Chem.* **277**, 36748–36754 (2002).
- Kim, S. H., Wang, W. & Kim, K. K. Dynamic and clustering model of bacterial chemotaxis receptors: structural basis for signaling and high sensitivity. *Proc. Natl Acad. Sci. USA* **99**, 11611–11615 (2002).
- Kim, K. K., Yokota, H. & Kim, S. H. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature* **400**, 787–792 (1999).
- Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
- Sourjik, V. & Berg, H. C. Binding of the *Escherichia coli* response regulator CheY to its target measured *in vivo* by fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA* **99**, 12669–12674 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. S. Parkinson for providing some of the plasmids, strains and antibody used in this study, and R. C. Stewart for providing antibody. We also thank D. Bray, K. A. Fahrner, J. J. Falke, J. S. Parkinson, T. Shimizu and A. Vaknin for comments on the manuscript. We thank P. Zucchi for technical help. This research was supported by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.C.B. (hberg@biosun.harvard.edu).

Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria

Omita A. Trivedi, Pooja Arora, Vijayalakshmi Sridharan, Rashmi Tickoo, Debasisa Mohanty & Rajesh S. Gokhale

National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India

The metabolic repertoire in nature is augmented by generating hybrid metabolites from a limited set of gene products^{1–3}. In mycobacteria, several unique complex lipids are produced by the combined action of fatty acid synthases and polyketide synthases (PKSs)^{4–6}, although it is not clear how the covalently sequestered

biosynthetic intermediates are transferred from one enzymatic complex to another. Here we show that some of the 36 annotated *fadD* genes, located adjacent to the PKS genes in the *Mycobacterium tuberculosis* genome, constitute a new class of long-chain fatty acyl-AMP ligases (FAALs). These proteins activate long-chain fatty acids as acyl-adenylates, which are then transferred to the multifunctional PKSs for further chain extension. This mode of activation and transfer of fatty acids is contrary to the previously described universal mechanism involving the formation of acyl-coenzyme A thioesters. Similar mechanisms may operate in the biosynthesis of other lipid-containing metabolites and could have implications in engineering novel hybrid products.

Analysis of the mycobacterial genome has revealed an array of genes involved in the lipid metabolism, including a large family of *fadD* genes⁷. These mycobacterial FadD proteins show homology to acyl-CoA synthetases that convert free fatty acids into acyl-coenzyme A (CoA) thioesters, the first step in fatty acid degradation. The functional importance of *fadD* genes in the biology of mycobacteria has recently been demonstrated by gene inactivation experiments^{8–11}. However, the precise biochemical functions of FadD proteins have not been investigated thus far. The computational analysis of mycobacterial FadD proteins revealed poor sequence homology with other acyl-activating enzymes such as firefly luciferases¹², adenylation domains of non-ribosomal peptide synthetases (NRPSs)¹³ and acetyl CoA synthetases¹⁴. The dendrogram obtained from multiple sequence alignments showed two distinct protein clusters (see Supplementary Fig. 1). The smaller cluster consisted of 12 FadD proteins with distinctly high sequence similarity. Interestingly, nine of these 12 *fadD* genes were located adjacent to PKS and NRPS gene clusters. In contrast, only two *fadD* genes (*fadD22* and *fadD10*) from the other cluster were in proximity with PKS and NRPS genes. On careful analysis of the FadD proteins

from the smaller cluster, it was observed that a number of conserved positions identified from multiple alignments corresponded with the active site residues of the adenylation domain of NRPS proteins. We therefore set out to investigate whether FadD proteins from the smaller cluster would activate fatty acids as acyl-adenylates and then transfer these acyl chains on to PKSs, in a mechanism analogous to the adenylation domains.

The small cluster includes the *fadD26* and *fadD28* gene products, which have been demonstrated to be essential for the biosynthesis of a virulent phthiocerol dimycocerosate lipid. It is interesting that despite the presence of other functional *fadD* genes and plentiful pools of acyl-CoA in the bacteria, the *fadD26* and *fadD28* mutants of mycobacteria lacked phthiocerol dimycocerosate on the cell envelope^{8–10,15}. The basis for this crucial non-redundant metabolic role was examined by analysing the enzymatic function of FadD26 and FadD28 proteins. Our computational analysis predicted two putative translational start sites for the FadD26 protein. We therefore cloned and expressed this protein from two independent start sites in *Escherichia coli*. The functional FadD26 protein was obtained from the downstream start codon. The enzymatic activity of the FadD26 and FadD28 proteins was monitored by using radio-labelled C₃ to C₁₈ fatty acids. Analysis of the reaction products using radio thin layer chromatography (TLC) showed identical retention factor (Rf) both in the presence and absence of reduced coenzyme A (CoASH) (Fig. 1a). The maximal activity for these proteins was obtained for long-chain fatty acids. These metabolites were then analysed on high-performance liquid chromatography (HPLC) (Fig. 1b, peak 1) and were characterized as fatty acyl-adenylate (AMP) by using electrospray ionization (ESI) mass spectrometry (MS) (Fig. 1c). The identity of the product was also confirmed by chemical synthesis of an authentic reference. This data establishes that both FadD26 and FadD28 proteins do not catalyse acyl-CoA formation as suggested in earlier studies^{8–10,15}, but instead synthesize

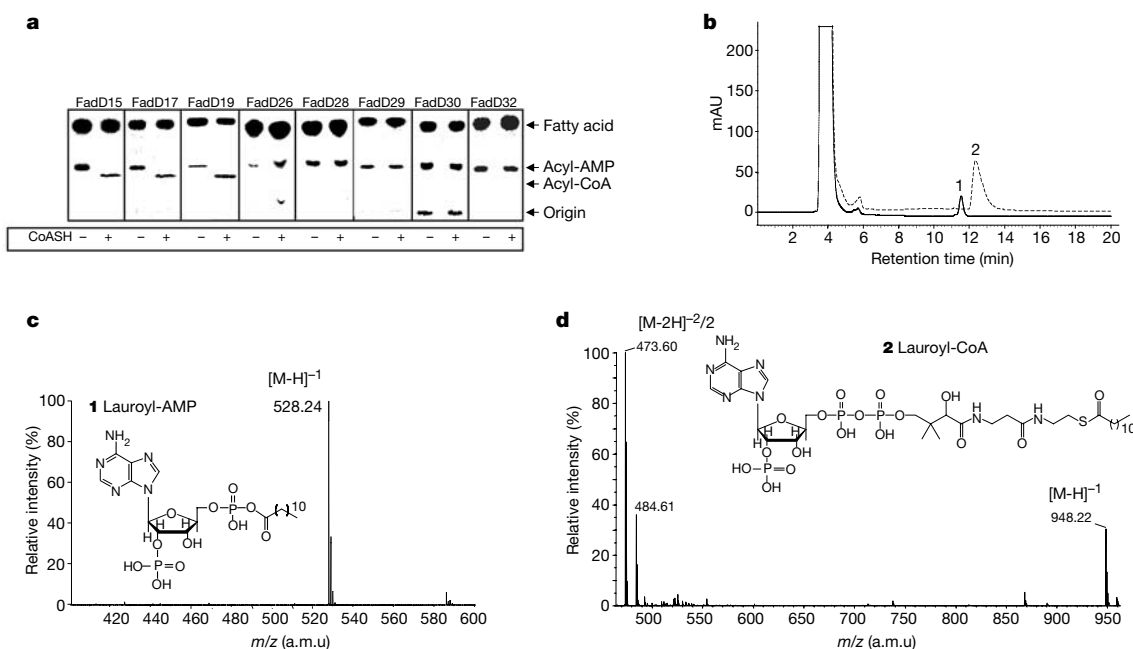


Figure 1 Chemical characterization of the FadD products. **a**, Radio-TLCs showing activity of various FadDs in the presence and absence of CoASH. The fatty acid migrates on the TLC with an Rf of 0.95. The acyl-AMP and acyl-CoA has an Rf value of 0.5 and 0.4 respectively. **b**, Reverse-phase HPLC chromatograms of reaction products synthesized by FadD28 (solid line) and FadD19 (dotted line) recorded at 259 nm. FadD28 catalysed the biosynthesis of lauroyl-AMP (structure **1**) while FadD19 formed

lauroyl-CoA (structure **2**). **c**, ESI-MS spectra for FadD28 product (Lauroyl-AMP). The mass/charge (m/z) peak at 528.24 corresponds to the molecular ion peak $[M-H]^{-1}$. **d**, ESI-MS spectra for the FadD19 product (Lauroyl-CoA). The peak at 948.22 corresponds to the molecular ion peak, whereas the peak at 473.60 corresponds to the $[M-2H]^{-2/2}$ ion.

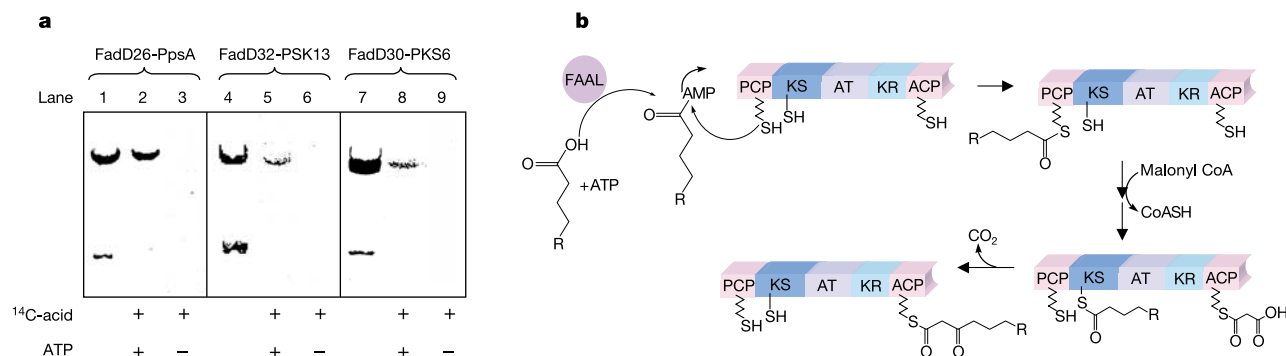


Figure 2 Activation and transfer of fatty acids on to the cognate PKS proteins. **a**, Autoradiogram showing radioactivity associated with PKS proteins. The Coomassie-stained PKS and FadD proteins are present in lanes 1, 4 and 7. Lanes 2, 5 and 8 show radio-labelled bands associated with PpsA, PKS13 and PKS6 proteins in the presence of their cognate FadD proteins and ATP. No radioactivity is observed in the absence of ATP (lanes 3, 6 and 9). **b**, The mechanism of fatty acid activation and transfer, mediated by FAAL proteins. The long-chain fatty acids are activated as acyl-adenylates and are

transferred on to the P-pant arm of the N-terminal PCP domain of PKS proteins. This acyl group is subsequently loaded at the active-site cysteine of the ketosynthase (KS) domain. The acyl transferase (AT) domain transfers the extender unit malonyl-CoA on to the ACP P-pant arm, which is used for a two-carbon chain extension through a decarboxylative condensation reaction. The ACP-bound intermediates are reduced by ketoreductase (KR) domain.

acyl-adenylates of long-chain fatty acids. In an earlier report, the purified FadD28 protein from *Mycobacterium bovis* BCG (which differs in 2/580 amino acids from *M. tuberculosis* FadD28 protein used in this study) was shown to catalyse formation of acyl-CoAs from their corresponding acids¹⁶. This may be because the metabolites were not chemically characterized and the enzymatic activity was inferred from radio-scintillation counting assays that cannot distinguish between the acyl-AMP and acyl-CoA. We therefore suggest that these two proteins should be annotated as FAALs.

Because the FadD26 and FadD28 proteins belong to the group of highly conserved 12 FadDs, we hypothesized that this small cluster might constitute a FAAL family of proteins. To test this hypothesis, three more *fadD* genes (*fadD29*, *fadD30* and *fadD32*) from this cluster were cloned and expressed in *E. coli*. As controls, three *fadD* genes from the larger cluster (*fadD15*, *fadD17* and *fadD19*) were also expressed in the heterologous host. All six of these purified FadD proteins catalysed the formation of acyl-adenylate from long-chain fatty acids in the absence of CoASH. The addition of CoASH to the

reaction mixture did not alter the product formation with FadD29, FadD30 and FadD32 (Fig. 1a). In contrast, a new product with lower Rf was observed for FadD15, FadD17 and FadD19 proteins (Fig. 1a), which also eluted at a different time on a reverse-phase HPLC (Fig. 1b, peak 2). This new product was established to be fatty acyl-CoA by using ESI-MS (Fig. 1d).

These studies corroborate our earlier computational analysis of the presence of two distinct groups of FadD proteins, one of which activates metabolic carboxylates as acyl-adenylates, while the others activate them as acyl-CoAs. The CoA analogues are used by the host in lipid degradation as well as for providing precursors in the biosynthesis of metabolites. We have recently characterized a new family of type III PKSs from *M. tuberculosis*, which uses such long-chain acyl-CoA substrates¹⁷.

To investigate the ability of acyl-AMP ligases to transfer the activated fatty acids specifically to the cognate PKSs, we cloned and overexpressed three PKS genes, *ppsA*, *pkS13* and *pkS6* in *E. coli*. The domain organization, as predicted by PKSDB^{18,19}, indicated the presence of two acyl carrier protein (ACP)-like domains in these

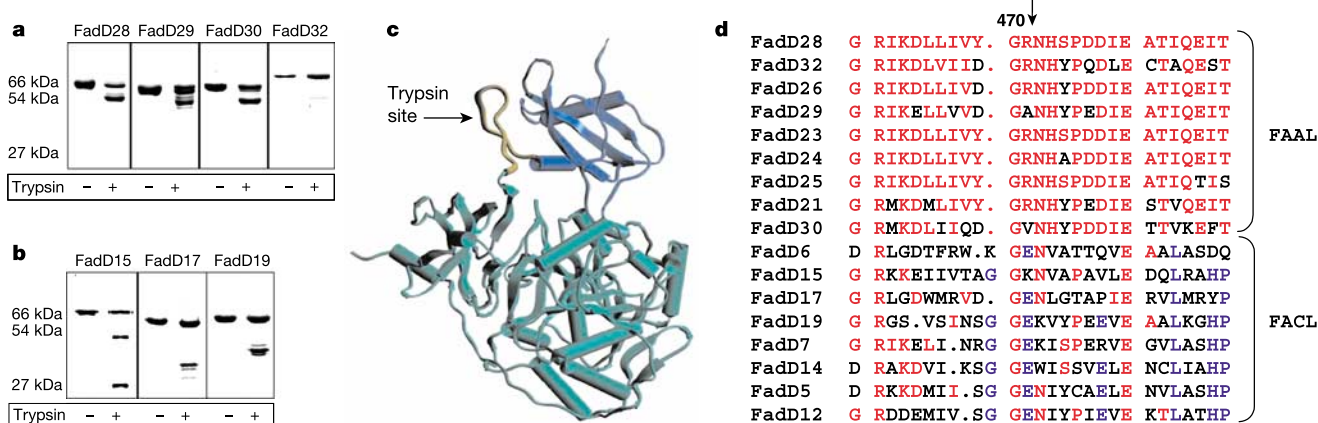


Figure 3 Structural analysis of FAAL and FACL proteins. **a**, Limited proteolysis of FAAL proteins with trypsin. **b**, Limited trypsin cleavage of FACL proteins. Each panel in **a** and **b** shows two lanes for each FadD protein, when incubated in the absence and the presence of trypsin. **c**, Three-dimensional model for FAAL proteins. The larger N-terminal domain is shown in green and the smaller C-terminal domain is shown in blue. The flexible hinge

region is drawn in yellow. The site of trypsin cleavage is indicated by an arrow. **d**, Multiple sequence alignment of the hinge region of representative members of FAAL and FACL proteins. The trypsin cleavage site on FadD28 is indicated by an arrow. Invariant residues are shown in red. The conserved amino-acid residues of FACL proteins are shown in blue.

PKS proteins. The amino-terminal ACP-like domain also showed homology to the peptidyl carrier protein (PCP) domain of NRPS proteins. The ACP and PCP are small homologous domains present in the multifunctional PKSs and NRPSs respectively, which are post-translationally modified at the active site serine residue. This phosphopantetheinylation (P-pant) modification of both PCP and ACP domains was achieved by co-expressing surfactin P-pant transferase (*sfp*) in *E. coli* host^{20,21}. The loading of activated acyl chains on to the N-terminal PCP domains of the cognate PKS proteins was monitored by incubating the proximal FadD–PKS protein pairs (FadD26–PpsA, FadD32–PKS13 and FadD30–PKS6) along with ¹⁴C-long-chain fatty acids and ATP. The proteins were resolved on SDS–polyacrylamide gel electrophoresis (PAGE) and analysed using autoradiography. Radioactive bands were associated with the PKS proteins in the presence (Fig. 2a; lanes 2, 5 and 8), but not in the absence (Fig. 2a; lanes 3, 6 and 9) of ATP. The transfer of fatty acids on to the PKS proteins and its subsequent extension is illustrated in Fig. 2b. This biosynthetic mechanism of incorporation of long-chain fatty acids is unprecedented and is reminiscent of adenylation domains of NRPS modules (see also Supplementary Fig. 2). Several lipopeptide biosynthetic gene clusters also contain open reading frames (ORFs) homologous to *fadD*^{22,23} and it is plausible that activation of lipid moiety might proceed through acyl-adenylate formation. Indeed, the loading domain of rifamycin synthetase initiates biosynthesis in a similar manner²⁴.

Structural investigations of the adenylation-forming family of enzymes have revealed conformational flexibility in the relative orientations of their two major domains^{12–14}. Our modelling studies of mycobacterial FadD proteins suggested that these proteins would be composed of a large N-terminal domain and a smaller carboxy-terminal domain. Because mycobacterial FadD proteins could be classified in two distinct families of FAAL and fatty acyl-CoA ligase (FACL) proteins, we examined whether they would exhibit different conformations by using limited proteolysis²⁵. Despite several potential tryptic cleavage sites, trypsin makes a single cut on the FAAL proteins (Fig. 3a). The smaller fragment of approximately 12 kDa was purified and characterized using mass spectrometry and N-terminal sequencing. The proteolytic cleavage site was mapped to a region between 470 and 471 residues of the FadD28 protein, which is predicted to be present in linker region between the two structural subdomains (Fig. 3c). Interestingly, this tryptic digestion pattern is similar to the adenylation domain of tyrocidine synthetase²⁶. The FACL proteins in contrast were much more susceptible to trypsin cleavage, and did not yield any particularly stable protein fragment (Fig. 3b). The comparison of the sequences of the hinge region between FAAL and FACL proteins also reveals differences in conserved amino-acid residues (Fig. 3d), suggesting that orientation of the C-terminal domain may be different between these two protein families.

The results presented here identify novel biochemical functions for the FadD proteins of mycobacteria and dissect mycobacterial carboxylic acid activation metabolism revealing new potential drug targets. More generally, the identification of a new class of FAALs and their role in the biosynthesis of complex hybrid metabolites also reveals a previously unknown molecular mechanism with which a limited number of genes can generate metabolite diversity. □

Methods

Materials

The bacterial artificial chromosome (BAC) library of *M. tuberculosis* H37Rv was obtained from the Pasteur Research Institute, France. Radioactive materials were purchased from Amersham Biosciences, Perkin Elmer Life Sciences and American Radiolabelled Chemicals, Inc. All other chemicals used were of analytical grade.

Cloning, expression and purification

M. tuberculosis *fadD* and *pks* genes were cloned by PCR amplification using gene-specific

primers from the BAC genomic library. The two ORFs of *fadD26* gene were of 1,881 and 1,752 base pairs (bp), which correspond to the start-codon positions of 3,243,697 bp and 3,243,568 bp respectively on the *M. tuberculosis* H37Rv genome. To minimize polymerase chain reaction (PCR) errors in larger PKS genes, shotgun libraries of BAC genomic DNA were constructed. The *fadD* and *pks* genes were expressed as hexahistidine-tagged proteins in the BL21-DE3 strain of *E. coli* using pET vector systems. Proteins were obtained in the soluble fraction by inducing cultures at low temperatures and isopropyl β-D-thiogalactopyranoside concentrations. These proteins were purified by Ni²⁺-NTA affinity chromatography followed by anion exchange chromatography using AKTA (Pharmacia) chromatographic system (Supplementary Fig. 3).

Enzyme assays

Fatty acid activation by various FadD proteins was analysed by radio-TLC and HPLC. The reaction mixtures included 50 mM Tris-HCl (pH 8.0), 100 μM ¹⁴C-fatty acid, 2 mM ATP, 8 mM MgCl₂ and 10 μM proteins in a 15 μl reaction volume and, wherever applicable, CoASH concentration was 2 mM. After incubation at 30 °C for 5 min, the reactions were quenched with 5% acetic acid and directly spotted on silica gel TLC plates. Products were resolved on TLCs in n-butanol/acetic acid/water (80:25:40) solvent system at 4 °C. The radioactive bands on TLC were detected using the phosphorimager (Fuji BAS5000). The reactions were scaled up for HPLC (Shimadzu) analysis. After incubation, the reactions were quenched and the products resolved on reverse phase (C8) column (Phenomenex) using an isocratic gradient at 50% B. Solvent A was composed of 20 mM ammonium acetate (pH 5.4) and solvent B was acetonitrile-methanol (85:15, v/v). Products were analysed using ESI-MS (API QSTAR Pulsar i MS/MS, Applied Biosystems).

Synthesis of fatty acyl-AMP standard

Fatty acyl-AMP was synthesized by modifying the procedure used in ref. 27. The acyl-AMP was characterized using ESI-MS.

Transfer of activated fatty acids to PKS proteins

FadD and its cognate PKS proteins were incubated along with 2 mM ATP, 8 mM MgCl₂ and 100 μM ¹⁴C-fatty acid at 30 °C for 15 min. Reactions were quenched and proteins were separated using SDS–PAGE. The gels were stained with Coomassie brilliant blue R250 and soaked in Amplify (Amersham) for 15 min. These were dried and exposed to phosphorimager. In the control reactions ATP was excluded from the reaction mixture.

Partial trypsin digestion

Various FadD proteins were incubated with 0.25 mg ml^{−1} or 0.05 mg ml^{−1} of trypsin at 30 °C for different time periods. The reactions were quenched and analysed on 12% SDS–PAGE gels.

Computer analysis

Primary sequences of the genes were obtained from TubercuList (<http://genolist.pasteur.fr/TubercuList/>). Multiple sequence alignments of FadD proteins were carried out by using the SeqWeb (<http://www.gcg.com>). The INSIGHT II package was used for depiction of structural models.

Received 29 November 2003; accepted 2 February 2004; doi:10.1038/nature02384.

- Moore, B. S. & Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **19**, 70–99 (2002).
- Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2498 (1997).
- Spaink, H. P. et al. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* **354**, 125–130 (1991).
- Gokhale, R. S. & Tuteja, D. in *Biotechnology* (eds Rehm, H.-J. & Reed, G.) 341–372 (Wiley, Weinheim, 2001).
- Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24**, 263–270 (1997).
- Sirakova, T. D., Thirumala, A. K., Dubey, V. S., Sprecher, H. & Kolattukudy, P. E. The *Mycobacterium tuberculosis* *pks2* gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276**, 16833–16839 (2001).
- Cole, S. T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
- Azad, A. K., Sirakova, T. D., Fernandes, N. D. & Kolattukudy, P. E. Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J. Biol. Chem.* **272**, 16741–16745 (1997).
- Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R. Jr Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature* **402**, 79–83 (1999).
- Camacho, L. R. et al. Analysis of the phthiocerol dimycoserate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J. Biol. Chem.* **276**, 19845–19854 (2001).
- Rindi, L. et al. Involvement of the *fadD33* gene in the growth of *Mycobacterium tuberculosis* in the liver of BALB/c mice. *Microbiol.* **148**, 3873–3880 (2002).
- Conti, E., Franks, N. P. & Brick, P. Crystal structure of firefly luciferase throws light on a superfamily of adenylation-forming enzymes. *Structure* **4**, 287–298 (1996).
- May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl Acad. Sci. USA* **99**, 12120–12125 (2002).
- Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M. & Escalante-Semerena, J. C. The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* **42**, 2866–2873 (2003).

15. Fitzmaurice, A. M. & Kolattukudy, P. E. An acyl-CoA synthase (*acoas*) gene adjacent to the mycocerosic acid synthase (*mas*) locus is necessary for mycocerosyl lipid synthesis in *Mycobacterium tuberculosis* var. *bovis* BCG. *J. Biol. Chem.* **273**, 8033–8039 (1998).
16. Fitzmaurice, A. M. & Kolattukudy, P. E. Open reading frame 3, which is adjacent to the mycocerosic acid synthase gene, is expressed as an acyl coenzyme A synthase in *Mycobacterium bovis* BCG. *J. Bacteriol.* **179**, 2608–2615 (1997).
17. Saxena, P., Yadav, G., Mohanty, D. & Gokhale, R. S. A new family of type III polyketide synthases in *Mycobacterium tuberculosis*. *J. Biol. Chem.* **278**, 44780–44790 (2003).
18. Yadav, G., Gokhale, R. S. & Mohanty, D. SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* **31**, 3654–3658 (2003).
19. Yadav, G., Gokhale, R. S. & Mohanty, D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335–363 (2003).
20. Lambalot, R. H. *et al.* A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
21. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
22. Duitman, E. H. *et al.* The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl Acad. Sci. USA* **96**, 13294–13299 (1999).
23. Tosato, V., Albertini, A. M., Zotti, M., Sonda, S. & Bruschi, C. V. Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. *Microbiology* **143**, 3443–3450 (1997).
24. Admiraal, S. J., Walsh, C. T. & Khosla, C. The loading module of rifamycin synthetase is an adenylation-thiolation didomain with substrate tolerance for substituted benzoates. *Biochemistry* **40**, 6116–6123 (2001).
25. Polverino de Laureto, P. *et al.* Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J. Mol. Biol.* **334**, 129–141 (2003).
26. Dieckmann, R., Pavela-Vrancic, M., von Dohren, H. & Kleinkauf, H. Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase 1. *J. Mol. Biol.* **288**, 129–140 (1999).
27. Keller, U., Kleinkauf, H. & Zocher, R. 4-Methyl-3-hydroxyanthranilic acid activating enzyme from actinomycin-producing *Streptomyces chrysomallus*. *Biochemistry* **23**, 1479–1484 (1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank S. K. Basu for discussions. P.A. is a Senior Research Fellow of CSIR, India. R.S.G. is a Wellcome Trust International Senior Research Fellow in India. This work was also supported by grants to the National Institute of Immunology by DBT, India.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S.G. (rsg@nii.res.in).

corrigendum

Delta-promoted filopodia mediate long-range lateral inhibition in *Drosophila*

Cyrille de Jossineau, Jonathan Soulé, Marianne Martin, Christelle Anguille, Philippe Montcourrier & Daniel Alexandre

Nature **426**, 555–559 (2003).

In the authors' published address, 'UMR 5539' should read 'CNRS UMR 5539'. An additional affiliation for the first and last authors (C.d.J. and D.A.) was omitted, who are also at the Laboratoire de Neurogénétique, Unité INSERM 432, Université Montpellier II. □

erratum

Folding proteins in fatal ways

Dennis J. Selkoe

Nature **426**, 900–904 (2003).

In this Insight Review Article, a competing interest was declared by the author that was accidentally omitted. This statement is: "D.J.S. is a founding scientist of Athena Neurosciences, now Elan plc." □

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejp.com

naturejobs

The fame game

The closest most scientists come to celebrity is conducting a conference presentation and perhaps having their image projected onto a big screen for the benefit of their peers in the back row. Researchers, even Nobel laureates, are relatively anonymous to the general public. But on 21 March, a group of scientists appeared on the cover of a glossy Sunday magazine read in millions of US households.

The credit for getting them onto the cover of *Parade*, in a space normally reserved for athletes, actors and pop singers, belongs to Research!America, a science-advocacy group in Alexandria, Virginia. But this public-relations coup has implications beyond a brief spell in the limelight. If non-scientists could be convinced to see researchers as rock stars, maybe scientists' careers would improve.

For example, Tshaka Cunningham, a graduate student at Rockefeller University in New York, sometimes fantasizes that if scientists were celebrities then he might have similar earnings to professional basketball players. And having a captive audience watch him clone a gene ought to make it easier for him to secure funding for his HIV research (see page 450).

There have been some isolated incidents where scientists have entertained such audiences. A few years ago, journalists queued up to have their copies of *Science* and *Nature* signed by the authors of both the public and private human-genome papers. And more recently, the world watched live as the Mars lander team awaited word that the probe had reached the surface safely.

Of course, scientists as celebrities could go too far — the trashing of green-rooms at conferences would be unfortunate. But some more panache with both peers and the public would be welcome. So scientists might consider adopting some Mick Jagger swagger when they fire up that next PowerPoint presentation.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

A fresh approach for
Spanish science

p448

CAREER VIEW

Recruiters & Industry

Back to the future
Graduate Journal
Paris, here I come...

Movers

Ali Raza

p450

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Renaissance in Spain



Sunny outlook: the future of science in Barcelona looks as bright as the view from Parc Guell.

Barcelona's cultural attractions, such as Gaudí's vibrant architecture, are well known. But it was the Catalan capital's new scientific opportunities that brought Madrid-born molecular plant biologist Soraya Pelaz back to Spain after five years at the University of California, San Diego. Her new place at the Institute of Molecular Biology in Barcelona is comparable with her well-funded lab in California, she says.

It's a major change. Science was stifled during Spain's 36 years under General Ferdinand Franco, only returning to life after the dictator's death in 1975. Since then, Spain has undergone dramatic changes, politically and socially, but in science it never quite managed to make up for the lack of infrastructure.

Widespread nepotism, a succession of governments uninterested in research (except for military uses) and a diehard bureaucratic apparatus have prevented Spanish science from becoming more competitive.

The country's best young scientists migrated in droves. But during the past few years science has climbed up the political agenda.

The trend is most notable in Catalonia, where a science-friendly regional government wants to turn Barcelona into a Mediterranean science showcase. A splendid postmodern glass building by Catalan architect Ricardo Bofill hosts the Institute for Photonic Sciences (ICFO), where scientists develop laser technology for use in research and medicine. By 2006, scientific staff will have grown from 20 to 150.

CORBIS

Reasons to return

Soraya Pelaz, who investigates flower development at the Institute of Molecular Biology in Barcelona, and Jorge Russo, a high-energy physicist at the University of Barcelona, have one thing in common. Both are group leaders with permanent contracts. Unlike most professors in Spain, they are not civil servants.

They are among 85 scientists hired by the Catalan Institution for Research and

Advanced Studies (ICREA), set up by the Catalan government in 2001 to bring high-profile researchers to the region. To qualify, scientists need a strong publication record and at least three years' postdoc experience abroad — and their expertise must be needed in Catalonia. Some 25 researchers a year win these tenured positions. Free of teaching obligations, they earn at least as much as university professors.

Another effort to stop the loss of talent from Spain, the Ramon y Cajal programme, was aimed at repatriating about 2,000 Spanish postdocs over the past three years.

For foreign scientists, ICREA offers a rare chance to enter the hermetic Spanish science system, with its rigid academic requirements. "Without ICREA," says Russo, "it would have been impossible for me to come."

Q.S.

"We are looking for senior researchers, young scientists and PhD students in physics, engineering and in the life sciences," says Lluís Torner, the ICFO's energetic director. The institute already has group leaders from Britain, Germany and Russia, and postdocs from France, Ireland and Spain. This global mix shows progress in a land that once could not keep its own talent. Scientists willing to leave a stable situation for Spain add a vote of confidence.

"You do pay a price if you stop everything and start from scratch in an empty lab," admits physicist Majid Ebrahim-Zadeh, who came to the ICFO last October from the University of St Andrews in Scotland. But with ICFO support he can build a group in experimental nonlinear optics without worrying about funding.

Palacios, head of research at Barcelona-based Almirall Prodesfarma, Spain's largest and most research-intensive pharmaceutical company. Almirall's research department for combinatorial chemistry is located in the Barcelona Science Park, opened in 2001. Here academic and industrial research groups share space, services and technology.

Meanwhile, in Barcelona's Olympic district at the opposite end of town is Miguel Beato, director of the Centre for Genomic Regulation (CRG), which was established in 2000. Beato hopes to emulate the European Molecular Biology Laboratory as a centre for independent research and training in molecular biology, bringing in young, energetic scientists from around the world. "But it is not easy to convince the best people to come to Spain," Beato says. "The next round of recruitment will be a test to see whether we already have the reputation it takes." The train bombings in Madrid earlier this month may complicate that test.

REINVENTING RECRUITMENT

Regional science initiatives are also under way in the Basque country, Castilla y Leon and Andalusia. But these sometimes clash with the central government's plans. Rolf Tarrach, a theoretical physicist at the University of Barcelona, resigned last year as president of Spain's Higher Scientific Research Council (CSIC) because he was frustrated about the administrative inflexibility, such as regulations that make it hard to bring in scientists who have PhDs earned outside Spain. To avoid the bureaucratic bottlenecks that keep many research institutes down, national science policies have begun exploring new avenues.

"The ability to recruit staff at any time was a condition for me to come back to Spain," says Mariano Barbacid, co-discoverer in the United States of the first human oncogene, and now director of the Spanish National Cancer Centre in Madrid. This centre, set up in 1998, combines basic cancer research with molecular diagnostics and target-based drug discovery. Barbacid manages this flagship of Spanish biomedical research in a US style, always looking for new funding opportunities. Four new research groups will be financed by a Spanish savings bank, for example. Eight groups will be launched in the next two years — an opportunity for foreign-trained Spanish postdocs.

A centre for cardiovascular research (CNIC) is planned nearby. Like its neighbour, the CNIC is a foundation of the Spanish health ministry, which finances its infrastructure and parts of the operating costs, but leaves all groups to raise their own research money through grants and collaborations. By 2007 it is due to have more than 300 scientists, including up to 15 internationally recognized group leaders.

Honduran Salvador Moncada, of the Wolfson Institute for Biomedical Research in London, is the CNIC's executive consultant. He is open to fresh research ideas. "Quality first, then definition of area," he says.

Investment in quality infrastructure for Spanish science could help to define the country as a scientific showcase. But to ensure that showcase is filled, the government must also build a system that can quickly and easily convert foreign PhDs — whether earned by Spanish citizens or expatriates — into credentials recognized by Spanish research institutions.

Quirin Schiermeier is *Nature's* German correspondent.



High ideals: construction work on Barcelona's Sagrada Família church.

Caitriona Creely, a young Irish biophysicist who in January joined the ICFO's biophotonics group, says that Barcelona now reminds her of Dublin in the 1990s, when the Irish government started pouring investment into science (see *Naturejobs* 4–5; 21 March 2002).

Hopes for a similar upswing in Spain rest mainly on the growth of the life sciences — and with several new institutes coming on line, this looks possible. The University of Navarra's Centre for Applied Medical Research will move into a new facility in June. A centre for research into biomedicine and tissue and organ transplantation is being built in Valencia. And researchers moved into the Basque institute of clinical research in Pais Vasco last year.

These and other life-sciences facilities in Spain are looking for screening experts, organic and medical chemists, and pharmacologists, says José-Maria

Web links

Catalan Institution for Research and Advanced Studies

♦ www.icrea.es

Institute of Biomedical Research of Barcelona

♦ www.pcb.ub.es/homePCB/live/en/p279.asp

Centre for Genomic Regulation

♦ www.crg.es

Spanish National Cancer Centre

♦ www.cnio.es/ing/index.html

GRADUATE JOURNAL

Paris, here I come...

What do basketball players such as Allen Iverson and LeBron James have in common with me? We're young African-American males living and working in the United States. How do we differ? They shoot jump shots and slamdunks in the US National Basketball Association, I do HIV research.

Also, unlike me, they are multimillionaires. As I scratch my head at sports salaries, I can't help but think that something must be out of balance.

Perhaps I'm alone in this view. Yet it seems that until crowds of adoring fans flock to our labs to root for us as we do our experiments, and we are greeted by dancing cheerleaders and thunderous applause each time we clone a vital gene, budding postdocs like myself can expect only minimal financial compensation. This harsh reality has affected my outlook and helped me to decide whether to do my postdoc in France or here in the United States (see *Nature* **427**, 762; 2004).

I reckon that, if I'm going to have a low income anyway, I may as well take the opportunity to experience a different culture and live in a foreign country. After all, many say that working abroad is an invaluable experience. I am hopeful that my time in Paris will prove this to be true. Otherwise, I may have to put down my pipettor and start working on my vertical leap. ■

Tshaka Cunningham is a fifth-year graduate student at Rockefeller University in New York.

RECRUITERS & INDUSTRY

Back to the future

Hundreds of years ago, medical scientists were broad-based and cross-disciplinary. But in the mid-twentieth century, they became specialized. Postgraduates spent years studying single genes, proteins or molecules. Industry demanded scientists who identified themselves as chemists, biologists, cellular biologists or biochemists.

In the past decade, the pendulum has swung the other way. Automation, robots and high-throughput screening, combined with genomics, created a semi-industrialization of drug discovery — and enormous quantities of data.

Disciplines such as chemical and biological informatics arose to deal with the information flood.

With these disciplines came the rebirth of an essential skill set. Industry

again needs cross-disciplinary scientists. Today's professionals figure out how basic science and advanced technology fit together. The ball is in the court of scientists who can assimilate data from a variety of areas. We are moving towards systems biology and biosimulation to understand the complexity of human health.

Our ability to generate data has outstripped our capacity to study them. The challenge is how to use the data we generate. We need to learn ways to bring technologies to bear in a way that manages risk and reduces the cost of drug discovery. How do we translate opportunity to sort out the ineffectual compounds and push forward the winners? It is the balance of science and advanced technology that will determine where we find our answers.

Just a short while ago, if you could clone and

sequence that was a great skill set to be able to put on your CV. Today, the scope of what the scientist needs to bring to the table has broadened. A solid molecular background will serve you well but you need to take the broader view, be able to work in teams, and collaborate with downstream partners.

To prepare for tomorrow, the bar will be set even higher. In another ten years we will see an increase in computer-aided drug-design applications of structural biology and protein engineering. This will be necessary because animal models are very expensive. Those who can adapt to this new environment will not only be able to forward their own careers, but at the same time move medical research closer to meeting our medical needs. ■

Michael Jackson is senior vice-president for drug discovery at Johnson & Johnson Pharmaceutical R&D in New Jersey.

MOVERS Ali Raza, medical director, Renovo, Manchester, UK



As a PhD student, Ali Raza made two choices that had profound effects on his career.

First, he opted to pursue pharmaceutical development rather than academic research. Second, he decided that earning a medical degree would be the best pathway to that destination.

Those decisions were informed by experience and personal contact. While Raza was in graduate school, several exposures to the pharmaceutical industry taught him that an MD's voice

often carries more authority in pharmaceutical research than a PhD's — even though a PhD may be more knowledgeable about a particular drug or disease mechanism. That information triggered the subsequent “ten-year diversion” into his medical training.

And while in medical school, Raza also weighed up the merits of being a practising physician versus being a pharmaceutical scientist. In seeing patients, the rewards are immediate, but are also limited to interacting with a few hundred patients a year. In the pharmaceutical industry, the rewards — a successful treatment for disease — are delayed, and even uncertain. But bringing a successful drug into the market could potentially touch millions, Raza says.

His strategy paid off. Raza says that his MD/PhD research background helped him to speed AstraZeneca's anti-

cholesterol drug Crestor from initiation of clinical trials to approval in several countries within five years.

Now, he is looking forward to a new challenge in his old stamping ground of Manchester, where he trained as a biochemist. But there are new challenges for him: he will be working at a small biotech firm with under 100 employees and in an area — wound healing — where he has little experience.

But he is happy to have traded commuting between continents for a ten-mile drive to central Manchester. He jokes that he is pleased to be reunited with the city's curry mile — a row of Indian restaurants. And he hopes he can replicate his drug-industry success in the biotech sector in the city where his career began. “It would be fantastic to build a billion-dollar company in my home town,” Raza says. ■

CV

2003–04: Head of clinical sciences, AstraZeneca, Goteborg, Sweden

1997–2003: Executive global product director and global team leader, AstraZeneca, Wilmington, Delaware

1994–97: Head of the medical department, Zeneca Pharmaceuticals, Toronto, Canada

1990–94: Drug team leader, ICI Pharmaceuticals, Alderley Park, UK